Honesty in the Provision of Expert Services: The Effect of Naturalistic Framings and Participants' Professions

Marco Piatti
BBus (Econ), MRes (Econ)

Submitted in fulfilment of the requirements for

Doctor of Philosophy (Economics)

School of Economics and Finance,
Queensland University of Technology

Principal Supervisor: Prof. Uwe Dulleck
Associate Supervisor: Prof. Benno Torgler
Associate Supervisor: Dr. Markus Schaffner

2016

Abstract

This thesis analyses inefficiencies in the healthcare market, by conducting a series of credence goods experiments to investigate how a naturalistic or neutral framing affects participants' honesty. This work is divided into three studies that will address different aspects of the forces driving healthcare costs. In the first study of this dissertation, a credence goods laboratory experiment with two different framings is conducted, namely a naturalistic framing relating to a medical environment and a neutral framing to investigate behavioural differences of a standard participant pool. Overall, it was found that medical framing has a positive effect on honesty. The second study extends the analysis by examining the behaviour of medical professionals from a large general practitioners conference in Germany in an abstractly framed credence goods experiment. Comparing the behaviour of students to medical professionals, it was found that medical professionals behave, on average, more honestly than students. However, it was also found that medical professionals overtreat, significantly more than students. Finally, the last study complements the two previous studies and translates them into a larger framework by conducting laboratory experiments with future health, law, engineering and accounting professionals. It was investigated in an experimental setting whether the behaviour of future professionals is affected by a naturalistic framing that is relevant to their future profession. Overall, it was found that naturalistic framings affect the group for which they are relevant, more than the remaining groups.

Keywords

Artefactual field experiment, Credence goods, Environmental framing, Honesty, Laboratory experiments, Medical framing, Medical professionals, Naturalistic framings, Neutral framing, Supplier-induced demand, Tax compliance framing

Dedication

This work is dedicated to my mother, my uncle Leo, and my cousin Jasmine, who all passed away much too early. I miss you all dearly. Love Marco

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature: M. Bratt

Date: 14.04.2016

Acknowledgements

Firstly, I would like to thank my supervisors, Uwe Dulleck, Benno Torgler and Markus Schaffner, for their guidance, moral support and determination, all of which were invaluable in completing this thesis. In particular I want to say thanks to Markus for his support in teaching me how to program experiments in CORAL. Secondly, I want to say thanks to David Savage for the many discussions and brainstorming sessions we had together over the years but in particular, during the last few months. Without these, I probably would not be writing this section at the moment. Furthermore, I would like to say a special thank you to Ben, Juliana, Jürgen, Stephen, Jonas, Ann-Kathrin, Daniel, Naomi, Osei, Romain, Suzanne, Zili, Vlad, Azhar, Yang, Tony and Poli for the great atmosphere and friendship we maintain with each other. It made this time so much more enjoyable. I will greatly miss our coffee breaks.

Furthermore, without Benno Torgler's support more than seven years ago I would never have undertaken this journey. First, Benno employed me as a research assistant; this then developed into a student-supervisor relationship and the completion of a Master by Research Degree. Benno is now not only my mentor but also one of my closest friends. Thank you so much for your friendship and sometimes speaking "tacheles" with me when I needed it. It means a lot to me.

I would also like to say thanks to the Economics Department at the University of Innsbruck and in particular, Professor Rudolf Kerschbamer, for the great time I had while I was visiting. I also very much appreciate that I was able to use their experimental laboratory for my experiments.

I express my profound gratitude to my family, specifically, to my wife and best friend Padaranee for all her support and patience over the last four years. Thank you so much. I would also like to thank my father for all his support throughout my life and particularly over this PhD journey. Last but not least, I would like to say thanks to my mother, who unfortunately passed away much too early, for all her support throughout my life.

Financial support from the Australian Research Council (ARC) through grant number DP110103653 is gratefully acknowledged.

Finally, I would like to thank the School of Economics and Finance, the School of Business and Queensland University of Technology staff. In particular I would like to thank the former Head of School, Professor Michael Kidd, and the current Acting Head of School, Associate Professor Tommy Tang, for their help and support. I thank the academic and administration staff of the School of Economics and Finance for the support throughout my time in the University. I acknowledge Diane Kolomeitz for professional copy editing and proofreading advice according to the guidelines laid out in the university-endorsed guidelines and the Australian Standards for editing research theses.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

On a daily basis, we trust and rely on experts to conduct a service that we ourselves cannot perform. Car repairs, taxi rides in an unknown city, medical treatments and dentist visits are all examples of occasions where we trust an expert to deliver the appropriate service. Providing medical services is perhaps one of the most important examples of an industry that relies highly on experts services. We are aware of the spiralling healthcare cost around the globe; however, we have little understanding of the forces driving these costs. This work seeks to better understand the factors that are responsible for the increasing costs of health care, through behavioural economic laboratory experimentation. Specifically, this research investigates how experts and students behave in framed experiments using abstract and naturalistic framing, with participant pools containing students, medical and non-medical professionals.

The medical sector is not only important for the wellbeing of a nations population; it also attracts a substantial amount of money from public spending. For instance, total expenditure on health as a percentage of gross domestic product has increased for numerous countries around the globe. For example, as depicted in Figure 1.0.1, between 1995 and 2012 for countries such as Germany, Austria, Australia or the US, total expenditure on health as a percentage of gross domestic product has increased between 12% and 32% respectively.

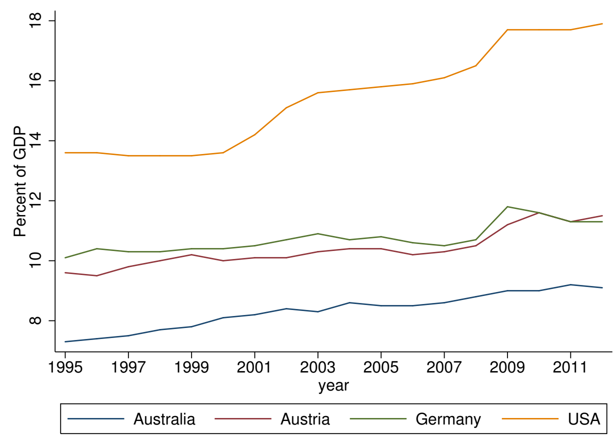


Figure 1.0.1: Total Expenditure on Health as a Percentage of GDP
Source: WHO (`http://apps.who.int/gho/data/node.main.75`,
accessed on 22.03.2015)

Given the substantial growth rate over the last decade raises the question on efficiency in the healthcare market, since over-prescription of pharmaceuticals (Wilson et al., 1995) and over-servicing by visiting medical officers in Victorian public hospitals[1] have already been documented in Australia. Furthermore, it can also be seen that in Austria and Germany, healthcare expenditure peaked between 2009 and 2010 respectively, and is since declining, which is in contrast to Australia and the United States. However, this is a misleading picture as both countries had a relatively high nominal GDP growth rate, as compared to the growth rate of health expenditure, which led to a decrease of total health expenditure as a share of GDP. In Austria, it decreased from 11.0% to 10.8% between 2010 and 2011[2] and in Germany, from 11.8% to 11.3% between 2009 and 2011[3] and was the actual reason for the reduction in total health expenditure as depicted in Figure 1.0.1.

Goods and services such as medical treatments, car repairs, and taxi rides

[1] `http://archive.audit.vic.gov.au/old/sr21/ags2101.htm`

[2] `http://www.statistik.at/web_en/statistics/health/health_expenditure/index.html` accessed on 6 June 2013.

[3] `https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2013/04/PD13_128_23611.html` accessed on 6 June 2013.

are all classified as credence goods in the economics literature. ‘Credence goods’ is the term that describes a situation in which the expert seller (he) knows more about the quality of a good that a customer (she) needs, than the customer herself (Dulleck and Kerschbamer, 2009). Such asymmetries of information between sellers and customers may give rise to inefficiencies. This thesis is analysing such inefficiencies in a credence good market by conducting a series of computerised economics experiments. In particular, this research is interested in honesty with regards to the delivery of health care services by medical professionals. As far as it is known, such experiments have never been done in a credence goods experimental setting with medical professionals as participants and a medical framing in the instructions. Extending the existing literature in this direction is important in increasing the external validity of these (highly application-relevant) other studies. For example, Ansink and Bouma (2013) state that “it is not clear whether and how subjects in the lab relate to, for instance, a ‘business’ or ‘Wall Street’ frame. From this perspective, framing should be matched to a participant pool to which the frame applied” (p. 3). Therefore, this dissertation will contribute to the understanding of differences between conventional laboratory experiments, and experiments with real world participant pools and real world framings.

This work is divided into three studies, which will address different aspects of the forces driving healthcare costs. The first study (Chapter 3), which is based on joint work with Uwe Dulleck, Rudolf Kerschbamer and Markus Schaffner, is entitled *The Experimental Economics of Credence Goods: The Effect of Medical Framing.* A credence goods laboratory experiment is conducted with two different framings, namely a naturalistic framing relating to a medical environment and a neutral framing to observe the inefficiencies prevalent in credence goods markets. For this study, a standard participant pool was invited. The novelty of this section is the fact that it is the first time that, in a credence goods experimental setting, a medical framing is used.

This study serves as a baseline to determine the effect of framing, that is, how decisions of a common participant pool are impacted when we introduce a naturalistic framing. Due to the naturalistic component of the framing, it might well be the case that the participants in the experiment alter their behaviour compared to the neutral framing, because participants can better emotionally connect to the task at hand. If this is the case, it would give a first indication if the (more morally connoted) medical context can mediate dishonest behaviour of experimental participants in a credence goods market. This provides us with a first benchmark for the studies introduced in later chapters.

The second study (Chapter 4), entitled *Medical Professionals as Participants in Credence Goods Experiments*, is also based on joint work with Uwe Dulleck, Rudolf Kerschbamer and Markus Schaffner. It extends the analysis in Chapter 3 by examining the behaviour of medical professionals. It is the first study that uses medical professionals from a large general practitioners' conference in Hamburg as participants in a credence good experimental setting, to examine one possible driving factor of the increasing healthcare cost around the globe. Supplier-induced demand (SID) or in other words, to overtreat patients, is often mentioned as a possible reason for the perpetual increase in healthcare expenditure. Due to the difficulties of measuring SID directly, most studies observe the frequency of revisits to a doctor and apply an empirical analysis. The approach in this study differs substantially to previous research investigating SID, by conducting a laboratory experiment instead of an analysis with observational data. This research is analysing the decision of medical professionals in an abstract-framed credence goods laboratory experiment. Utilising such an approach allows us to observe medical professionals' decisions, with respect to undertreatment, overcharging and our main variable of interest, overtreatment. It will therefore be of great interest to analyse results and see if those obtained with students in the previous study are similar to the results with real medical professionals or whether medical professionals behave differently i.e. more honestly.

The third study (Chapter 5), entitled *Naturalistic framings and honesty of future professionals: A triple experiment*, is joint work with Uwe Dulleck, Markus Schaffner and Benno Torgler. This study complements the two previous studies and translates them into a larger framework. The existence of ethical dilemmas regarding workplace incentives and professional practises or unwritten rules to overcome these dilemmas is not only a problem in the medical sector but also in other areas such as law. In the public eye, the image of professions such as doctors and engineers are rated very highly, while lawyers or politicians receive a less favourable assessment. It is interesting to observe if the perception of the public is accurate when we analyse the behaviour of different prospective professionals with respect to honesty in the experiment. Since the professions we are interested often occupy important positions within society, it is important to know if the people who self-select themselves as pursuing a career in one of these professions possess the required traits, such as honesty, to be successful. To analyse prospective professionals prior to their work experience is an advantage, as they have not yet learned the heuristics that could bias their behaviour in this experiment. There is no other study that has utilised a credence goods experiment with medical framing, a tax compliance experiment with a public goods structure and a self-reporting task with an environmental framing, all combined in one single experiment.

Overall this research finds that framing has a strong effect on participants in the experiment. In particular, I find that naturalistic framings affect the group for which they are relevant (e.g. medical framing for future health professionals) more than the remaining groups in the experiment. Furthermore, medical professionals exhibit a more pro-social behaviour compared to students in the experiment. Finally, it was found that accountants were the most honest future professional participants in our experiment, while lawyers were the least honest participants.

This work continues with a literature review (Chapter 2), followed by the three aforementioned studies and concludes (Chapter 6) with remarks as well as a discussion of the implication of the studies, provides some policy implication, emphasises some shortcomings and provides an outlook for future research opportunities.

Chapter 2

Literature Review

A large fraction of the current approach to behavioural economics encompasses laboratory experiments. The advantage of using laboratory experiments is the fact that this is a controlled environment where we can control a number of variables and isolate effects. This section elaborates on the experimental game used, and gives a brief explanation of SID, some specifics regarding laboratory experiments, the concept of framing in experiments, a very brief discussion on other regarding preferences, as well as the experimental software that was utilised.

2.1 Credence Goods

Darby and Karni (1973) were the first to introduce the term credence goods and added it to Nelson's (1970) classification of ordinary, search and experience goods. Ordinary goods (such as rice) have well-known characteristics, and people know where to obtain them. Search goods (like shoes) have to be examined before buying, in order to observe their characteristics. Experience goods (like wine) have unknown characteristics, but they are discovered after purchasing or using them. The term credence good describes a situation in which an expert knows more about the quality of a good a consumer needs than the consumer herself (Dulleck and Kerschbamer, 2009). One particular characteristics, that differentiates a credence goods from the before mentioned goods, is the fact that consumers cannot even be sure ex-post if they received the appropriate treatment or not. The economic literature provides both anecdotal and empirical evidence suggesting that *undertreatment*, *overtreatment* and *overcharging* are important problems in the provision of expert services in many countries and industries. In the medical context, Emons (1997) quotes a Swiss study stating that the average person's likelihood of receiving one of seven major surgical interventions is one-third above that of a physician or a member of a physician's family. This indicates that better informed and educated patients are less likely to be offered or accept unnecessary and expensive surgical interventions. Emons goes on to mention a study by the Federal Trade Commission that documents the tendency of optometrists to prescribe unnecessary treatments. Iizuka (2007) finds that doctors react to mark-up differences in the Japanese drug prescription market where doctors often prescribe and distribute medications. Additionally, he also finds that physicians prefer to prescribe cheaper drugs to patients. Gruber et al. (1999) show that the relative frequency of caesarean deliveries, compared to normal child births, responds to the fee differentials of health insurance programs. In support, Gruber and Owings (1996) state that the frequency of caesarean section deliveries is negatively correlated with birth rates. Hughes and Yule (1992) find that the number of cervical cytology treatments is positively correlated with the fee for this treatment. Jürges (2007) and Fuchs (1978) indicate that a large share of patients' demand for health care services is supply-driven; physician density has a significant positive effect on the number of doctor visits and operations. This confirms the previous statement that *overtreatment* in expert-supplied healthcare adds considerably to the overall expenses of the system.

Based on the standard assumption made in economics of own-money maximising preferences, these problems were addressed theoretically by several academic articles that aim to identify the conditions for the efficient provision of credence goods (see,e.g., Pitchik and Schotter, 1987; Wolinsky, 1993; Taylor, 1995; Glazer and McGuire, 1996; Emons, 1997, 2000; Sülzle and Wambach, 2005; Fong, 2005; Dulleck et al., 2011, 2014; Bester and Dahm, 2014; Dulleck et al., 2015). These conditions are organised in a comprehensive survey paper, published by Dulleck and Kerschbamer (2006). They describe markets for credence goods by at least three interrelated characteristics, according to their assumptions of the markets. They have to state (1) the technology of the suppliers; (2) the degree of competition combined with the structure of the market; and (3) the information structure of customer and courts. Additionally, their survey paper provides a unifying framework into which the various approaches and their predictions can be integrated (see Table 3 in Dulleck and Kerschbamer, 2006, p. 34). The model shows that two conditions are crucial for the efficient provision of credence goods:

1) **Liability (L)**

The expert is required, by law or some other institution, to provide a good that satisfies the customer's needs. Liability prevents the problem of *undertreatment*, but not necessarily the problems of *overtreatment* and *overcharging*.

2) **Verifiability (V)**

The customer is able to observe and verify the delivered product or service ex-post. A patient may, for some medical treatments, be able to verify the provided service, like taking an x-ray, while for others, like an operation under anaesthetic, this may not be the case. Verifiability solves the problem of *overcharging*, but not necessarily the problems of *over*- and *undertreatment*.

Standard theory predicts that without at least one liability or verifiability,, experts will always provide the lowest quality (*undertreatment*) and charge for the highest one (*overcharging*). This phenomenon leads to problems similar to those in Akerlof (1970)'s "*Market of Lemons*" model in which the market will break down due to the fact that goods traded in a market consisting of information asymmetry between a buyer and seller can degrade, hence, only leaving lemons behind.

In a recent paper, Dulleck et al. (2011) were the first to study the determinants for efficiency in a credence goods market by utilising a laboratory experiment. Their experiment was conducted in a neutral or abstract credence goods environment, i.e. an environment without any reference to the medical sector or another real world credence goods situation. Also, the participants in the experiment are students and not experts from the field. They observe that if liability holds, the credence goods problem disappears but verifiability has, at best, a minor effect. This empirical result is in contrast to the theoretical result predicted by Dulleck and Kerschbamer (2006), where they suggested that verifiability might be one way to overcome the credence goods dilemma. The results also demonstrated the limits of standard theory in explaining all observed behaviour and that some institutions, identified in theory to solve the credence goods problem, do not work in practice or at least in the laboratory. Apart from their project, Huck et al. (2007, 2012) study markets of experience goods and goods and services, which also involve problems of asymmetric information but differ because the consumer finds out, after purchasing, which quality she received. Huck et al. (2007, 2012) study the importance of trust in such markets. The results presented in Dulleck et al. (2011) and Kerschbamer et al. (2009) are complementary to the Huck et al. (2007, 2012)'s results and they are the very first experimental studies on behaviour in credence goods settings. Since then, a few more experimental studies for credence goods markets have emerged. Huck et al. (2014) analysed in an experimental setting how medical insurance and the free choice of a physician correlates with the overtreatment of patients, and Balafoutas et al. (2015a) investigated tax evasion in a credence goods laboratory experiment. A paper published by Beck et al. (2014) is the first that used an artefactual field experiment[1] in a credence goods experimental

[1] "An artefactual field experiment is the same as a conventional lab experiment but with a nonstandard participant pool"(Harrison and List, 2004, p. 1014). In this case car mechanics.

study to analyse the behaviour of car mechanics. They find that car mechanics overtreat significantly more than students. Balafoutas et al. (2013) is one of only a few field experiments aimed at analysing the determinants of fraud in a credence goods market. They investigated taxi drivers in the Greek capital, Athens. Their results indicate that *overtreatment* (taking a detour) was the main determinant of fraud, with 46% of passengers affected by it. *Overcharging* (manipulating fares) only accounted for 11% of fraudulent behaviour. *Undertreatment* (not reaching the requested destination) is not a problem in this type of credence goods market, as it is very easily observed. Another conducted field experiment is that of Schneider (2012), who investigated fraudulent behaviour in a credence goods market for car repairs. He found that *overtreatment* and *undertreatment* was of major concern. However, he goes on to state that incompetence of the car mechanic cannot be ruled out to have caused these results. The latest field experiment conducted in a credence goods market is Balafoutas et al. (2015b) who look at second-degree moral hazard in taxi rides in an unknown city. In their moral hazard treatment, passengers, in contrast to their control treatment, particularly state that they need a receipt *such that they can get reimbursed by their employer.* Their result suggests that passengers in that condition are significantly more likely to pay a higher price than passengers in the control treatment, where they only ask for a receipt. The second-degree moral hazard problem is also very relevant in medical situations where often insurance companies pay for the medical treatment costs, rather than the patients themselves. Furthermore, Rasch and Waibel (2015) analyse empirically the occurrence of overcharging in the market, for car repairs. They find that the higher number of competitors in the vicinity lowers the incentive to overcharge. This dissertation builds on the credence goods experimental research by using the same design as in previous credence goods laboratory experiments, however, this research extends the previous experiments by employing a naturalistic, specifically a medical framing in addition to the neutral framing. Moreover, the participant pool consists not only of students, but also of medical professionals from a general practitioners conference in Germany.

2.2 Supplier-Induced Demand (SID)

Arrow (1963) contends that the information asymmetry between a doctor (expert) and a patient (consumer) promotes the problem of SID within the healthcare sector. SID is the term used in health economics to describe 'services ordered by a physician for a patient, that the patient would refuse if he or she had the same medical knowledge and expertise as the physician' (Hay and Leahy, 1982). While most studies that analyse SID use an empirical approach (Fuchs, 1978; Gruber and Owings, 1996; Gruber et al., 1999; Jürges, 2007), this research deviates from such an approach and investigates the behaviour of medical professionals directly. Various reasons exist for the occurrence of SID in the healthcare market. Kessler and McClellan (1996) state that the fear of doctors being held liable for malpractice could be one reason why doctors choose to overtreat patients, just to be in a safer position. Bickerdyke et al. (2002) state two different reasons for the occurrence of SID, namely a physician's self-interest and the well-being of their patients. This thesis aims at better understanding the motives and incentives and the resulting behaviour of health care providers that cause such inefficiencies, by conducting an artefactual field experiment with medical professionals from a large conference in Germany.

2.3 Experiments

In economics, the utilisation of laboratory experiments has emerged relatively late when compared to other social sciences, such as psychology. Smith (1962) was one of the first who used experimental economics to test market behaviour in a classroom setting. Since then, experimental economics has come a long way. In particular, laboratory experiments are used to analyse coordination games, market games and auction games as well as social preferences.[2]

The majority of experiments conducted in the laboratory use university

[2] To analyse social preferences in different circumstances, games such as the dictator game, ultimatum game, trust game or public goods game are utilised

students as their participants, who are normally not experienced at the task at hand, nevertheless, for most theoretical experiments investigating decision-making in general that does not matter. On the other hand, for policy motivated experiments it is important to know how professionals behave compared to a general student population before implementing a policy change (see for example Hannan et al., 2002, where experienced MBA students behave differently to undergraduate students).

From psychology, we know that it is important how something is presented to participants in experiments, as this can influence their decisions. Ariely (2008, p.1) nicely shows an example for a subscription to *The Economist* where the framing or presentation of the different subscriptions options determines the selection that the marketing people at *The Economist* want people to choose. Hence, it is important that the participants in experiments behave as naturally as possible to avoid undesirable or biased results. Preventing such biased behaviour is not always entirely possible, however, it can be minimised. Employing a neutral framing, for example, can help to overcome such a problem, as participants cannot associate their behaviour in the experiment to a real world decision and therefore they should not behave unnaturally.

However, a neutral framing might not always be the best solution to observe participants' behaviour in laboratory experiments. For policy-related experiments, it might be important to have naturalistic framings relevant to the area of the workforce that is to be examined or for which the policy ought to be implemented. For experimental participants to be able to psychologically relate to the situation at hand might possibly enhance a pro-social behaviour compared to participants playing the same game but with a neutral framing. For instance, the free rider problem might be overcome with a naturalistic framing. A key sector where the behaviour of participants in the market plays an important part for society is the medical sector. In the public perception, medical professionals are more honest and ethical than people working in many other jobs. For example, interpreting the 2014 Gallup Poll question on honesty and ethics in different fields, it can be seen that jobs in the medical sector are rated very positively compared to professions such as bankers or lawyers.[3] But are medical professionals living up to their public perception? Interestingly, not many studies exist that use medical professionals or other workforce professionals of different areas to answer such questions. As it is difficult to observe the behaviour of such professionals directly, laboratory experiments with an abstract and naturalistic framing are a good way to overcome such obstacles.

Using real world participants is still not a very common practice in experimental economics, in particular in experimental studies that focus on health economics, and can therefore contribute to the discussion of external validity of laboratory experiments (see Harrison and List, 2004, for a discussion on the nature of the subject pool).

2.4 Framing in Experiments

The importance of framing, hence the style of presenting experimental results, has been indicated by a large fraction of the literature (Andreoni, 1995; Dufwenberg et al., 2011; Ellingsen et al., 2012; Gächter et al., 2009; Kühberger, 1998; Kühberger et al., 1999; Levin et al., 1998; Tversky and Kahneman, 1981, 1986; Tversky et al., 1990). Framing plays an important part in psychological and economics experiments. In their seminal paper, Tversky and Kahneman (1981) showed that preferences shift if the task that participants have to undertake is framed positively or negatively. Tversky and Kahneman (1986) point out that preferences between options should be independent of their representation, and therefore, not violating the invariance axiom of expected utility theories.

Such preference reversals were first mentioned by Lichtenstein and Slovic (1971). Kahneman and Tversky (1979) explain this shift with Prospect Theory. For more papers on preference reversal, see Lichtenstein and Slovic (1973); Grether and Plott (1979); Slovic and Lichtenstein (1983); Tversky et al. (1990) to name just a few. For a meta-analysis on Asian disease-like studies see Kühberger et al. (1999). In their survey paper, Levin et al. (1998) define fram-

[3] http://www.gallup.com/poll/1654/honesty-ethics-professions.aspx

ings that are concerned with providing information or describing situations in either a positive or a negative attribute, *valence framing*. Levin et al. (1998) go on to further break up valence framing into risky choice, attribute and goal framing. Dufwenberg et al. (2011) describe label framing as a situation by which participants are confronted with an equivalent task, same material incentives but different wording. The paper written by Andreoni (1995) is a good example of a paper that could be labelled as such. He conducted a public goods experiment with two different framings, one with a positive and the other with a negative framing. In the positive framing the participants' choice is framed as contributing positively to the public good, while in the negative framing, incentives are identical to the positive framing; the participants' choice is framed as buying a private good that, since the opportunity cost is the acquisition of the public good, leads to a negative outcome of the other participants. The result of the experiment showed that participants are cooperating much more in a public good experiment that is positively framed, compared to a public good experiment that is negatively framed, even though the incentives of the experiment have not changed. Andreoni (1995) explains this result with one hypothesis that could be called pure altruism but is highly unlikely or an alternative hypothesis, which he thinks is more likely, in that people get a warm-glow by behaving nicely to other people. However, to obtain such a result the strength of the warm glow needs to be greater than the cold prickle sensation of doing bad to others. At the same time, though, Andreoni (1995) states that more research needs to be undertaken to be able to affirm for certain that an asymmetry concerning positive and negative externalities can be generalised to other aspects of human interactions.

Deutsch (1958) in his seminal paper was the first to investigate trust by conducting an economics experiment. He showed that the choice behaviour of participants when playing a prisoner's dilemma game depended heavily on whether the instructions were framed to make participants feel individualistic or cooperative. This is in line with Ellingsen et al. (2012), who in their first experiment showed that the framing of different circumstances significantly affects the behaviour of participants in social dilemma situations. As Camerer (2003) expresses it, "There is little doubt that describing games differently can effect behaviour; the key step is figuring out what *general* principles (or theory of framing) can be abstracted from labelling effects" (p. 75). In their neutrally and medically framed experiment, Ahlert et al. (2008) found that economics students behave less selfishly in the medical framing compared to the neutral framing. On the other hand, medical students behaved more selfishly in the medical framing. Their explanation is that economics students appear to obtain cold feed in the unfamiliar medical framing, while for medical students, their professional norms appear to emerge more clearly in the familiar framing, hence external validity is of high importance as it could be the case that medical professionals are not sensitive to a medical frame. This would indicate that we cannot rely as much on moral suasion. Furthermore, Cooper et al. (1999) in their experiment conducted with students and managers in China, use two different framings. One is a neutral framing not related to the managers' profession while the other framing - they call it context framing - is related to the profession of the managers. They find that the context framing influenced the behaviour of the manager much more compared to the neutral framing. They relate this to the fact that managers could rely on past experience and similarities to their daily tasks at work. In Chapter 3 a medical framing and a neutral framing are used in the experiment conducted with student participants, while in Chapter 4 medical professionals are used as participants in a neutral framing environment.[4] As Al-Ubaydli and List (2013) point out, 'the best predictor of cheating rates in the natural field experiment is behaviour in sterile laboratory treatments with neutral language instructions' (p. 33). Hence, the results of this research are a valuable contribution to the credence goods, framing and artefactual field experiment literature.

[4]Unfortunately, I was not able conduct the medical framing experiment with medical professionals due to difficulties recruiting medical professionals for my experiment at the medical conference in Germany.

2.5 Other Regarding Preferences

Recent economic research shows that the behaviour of people, including experts, may be motivated by an intrinsic desire to solve their customer's problem (Frey, 1997) and/or by a preference for honesty (Brandts and Charness, 2003; Charness and Dufwenberg, 2006; Gneezy, 2005; Alger and Renault, 2007; Mazar and Ariely, 2006; Severinov and Deneckere, 2006); they also might have a desire for efficiency (Charness and Rabin, 2002) and/or a concern for how the seller's utility compares to the customer's (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Additionally, the behaviour of people could be influenced by aversion to guilt (Battigalli and Dufwenberg, 2007; Ellingsen et al., 2010). While those types of preferences are modelled theoretically and found to be relevant empirically, no research exists that answers the question whether and to what extent real world experts in the health sector are driven by such motives. Previous theoretical research (Kerschbamer et al., 2009, 2015) has developed a theory of behaviour of experts with such non-standard preferences. Their findings show that other-regarding preferences can dramatically change the predictions of provision behaviour of medical goods and credence goods in general. Eckel and Gintis (2010) state that other regarding preferences are driven by social norms, and social norms are environment specific. This is particular important for the experiments in Chapter 3 and 5, in which experiments with context specific environments such as a medical framing, an environmental framing and a tax compliance framing are included. Participants in before-mentioned experiments are perhaps behaving more honestly as they believe this is the socially appropriate behaviour.

2.6 Experimental Software

This section discusses the experimental software that was used to conduct all experimental sessions at the University of Innsbruck (*IBK*), a medical conference in Hamburg (Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin (*DEGAM* 2014)) and at the Queensland University of Technology (*QUT*).

For this thesis, a new experimental software named *CORAL* is used (Schaffner, 2013), which is a lightweight framework for experimental economic experiments. *CORAL* is a HTML Velocity and Java-based program that allows for a more up-to-date design of an experimental environment compared to previous experimental software, such as zTree developed by Fischbacher (2007). One of *CORAL*'s advantages is that it is an open-source software, which breaks down the experiment to a simple step-by-step execution of templates or script files (see Schaffner, 2013, for a detailed introduction of *CORAL*) or the homepage that was particularly created for new users[5].

[5] `https://github.com/mas802/coral-econ`

Chapter 3

The Experimental Economics of Credence Goods: The Effect of Medical Framing

3.1 Introduction

Framing plays an important part in economic and psychological decision making experiments, as it can alter a participant's behaviour in certain situations. In their seminal paper, Tversky and Kahneman (1981) have shown that participants shift their preferences if the task is either framed positively or negatively. Furthermore, Cooper et al. (1999) find that in an experiment conducted with students and managers in China, the context framing, which was relating to the managers' profession, influenced the behaviour of the managers much more compared to the neutral framing, while for students it did not matter much. They relate this to the fact that managers could rely on past experience and similarities to their daily tasks at work. Supporting their result is a recent study by Cohn et al. (2014), who find that banking professionals behave less honestly if the framing of the experiment relates to a financial environment compared to a neutral framing. This confirms the perception that the banking environment weakens and undermines the morally desired behaviour of honesty in employees. In this chapter, it is of interest whether participants behave differently when a medical framing to which they can relate is applied. The novelty of this chapter is the framing of the experiment. As far as it is known, this research is the first to apply a medical framing in a credence goods experiment. Additionally, in contrast to Dulleck et al. (2011), a strategy method is also applied in our experiment. It is found that in this baseline treatment, students undertreat, overtreat and overcharge significantly less ($p < 0.01$) in medical framing than in neutral framing. Specifically, it is observed that in medical framing, undertreatment is 45 percentage points less likely to occur than in neutral framing. However, when liability or verifiability is introduced, the effect of medical framing is much less apparent. The reminder of the paper is structured as follows: In Section 3.2 the methodology of the paper is discussed. Section 3.3 describes the experimental design. The results are presented in Section 3.4 and last, Section 3.5 concludes.

3.2 Model and Methodology

3.2.1 Baseline

A simplified version of the Dulleck and Kerschbamer (2006) credence goods model will be the commencing point of this study. In this game, there are two players: an expert seller S (he) and a consumer C (she). In a first step, the expert seller posts two prices, a price p^h for the high quality service and a price p^l for the low quality service, with $p^h \geq p^l$. The consumer gets to know these prices and decides if she wants to interact with the expert seller or not. If she decides not to interact in the market, no trade takes place and both players receive the outside option of $o \geq 0$. If the consumer decides to stay in the market, nature determines with probability h that she needs the high quality treatment and with probability $1-h$ the low quality one. The expert learns about the needs of the consumer and provides either the high quality treatment t^h or the low quality treatment t^l, where t^h incurs costs of c^h and t^l of c^l respectively, with $c^h \geq c^l$. In the last decision, the expert charges one of the prices he previously defined for the respective treatment. The customer receives a payoff $v > 0$ if the service provided by the expert was sufficient (*i.e.* if she needs the high quality and obtains the high quality, or she requires the low quality and either receives the high quality or the low quality) or *zero* otherwise. The customer has to pay the price of the service regardless of receiving the sufficient or insufficient quality provided by the expert. More formally, let $\theta \in \{l, h\}$ be the index of a customer's type of problem, $\delta \in \{l, h\}$ be the index for the quality of the treatment provided and let $\gamma \in \{l, h\}$ be the index of the quality of treatment actually charged for. The material payoffs of the expert seller π_S and consumer π_C are therefore as follows:

$$\pi_S(p^l, p^h, \delta, \gamma) \quad = \quad p^\gamma - c^\delta \tag{3.1}$$

$$\pi_C(p^l, p^h, \theta, \delta, \gamma) = \begin{cases} -p^\gamma & \text{if } \theta = h \text{ and } \delta = l, \\ v^\theta - p^\gamma & \text{otherwise.} \end{cases} \tag{3.2}$$

Figure 3.2.1 illustrates the sequence of decisions in this game and the payoffs provided to each player.

In the following, the basic setup is extended and some institutional restriction such as *liability* and *verifiability* is imposed on the expert sellers action space.

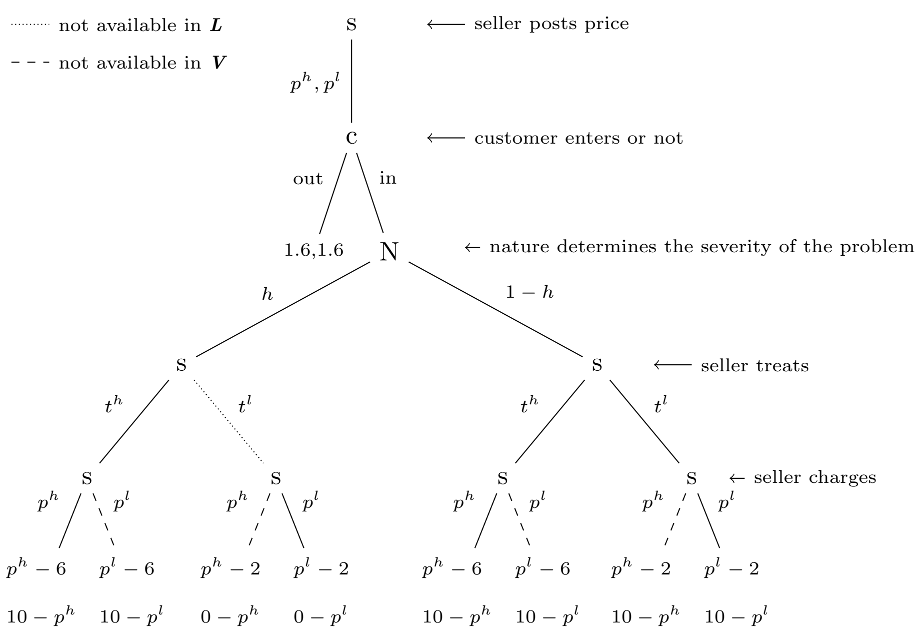


Figure 3.2.1: Game Tree

3.2.2 Adding Institutional Restrictions: *Liability* and *Verifiability*

The *liability* condition in credence goods markets infers that expert sellers always provide the required quality that solves the consumer's problem. *Liability* does therefore prevent the problem of *undertreatment*; it does not, however, preclude the problem of *overcharging* or *overtreatment*. Imposing the *verifiability* condition on the other hand, can prevent the problem of *overcharging* because consumers can observe and verify ex-post the quality provided by the seller (however, without knowing if it was the appropriate one), but it does not preclude the problems of *undertreatment* or *overtreatment*.

The two institutional restrictions *liability* and *verifiability* and all the combinations create four different sets of possible actions for the expert seller.

The four conditions are treatment **B** (*Baseline* where neither *liability* nor *verifiability* holds), treatment **L** (where *liability* but no *verifiability* holds), treatment **V** (where only *verifiability* but no *liability* holds), treatment **LV** (where *liability* and *verifiability* hold (no experiments were conducted in this treatment due to the fact that in the paper published by Dulleck et al. (2011) no additional information could be won from this treatment, because adding more institutional restrictions reduces fraud in the first place since it leaves participants with less choices. Furthermore, it also allowed us to maximise the number of participants in the remaining treatments.). These treatments are carried out similar to Dulleck et al. (2011) but this study varies the framing of the experiment from a neutral to a medical framing. Additionally, in contrast to Dulleck et al. (2011) a strategy method is also used in all the experiments. To be exact, only a strategy method for the player that was allocated the role of the seller or doctor was implemented. Hence, we obtained for every round an observation on the decision made by the expert seller on the treatments he would have provided and the price chosen, even for the cases when the consumer decided not to participate in the market.

3.3 Experimental Design

In the medical framing the same experimental design is followed as in the neutral framing (see Figure 3.2.1 (Game Tree) in previous section), however, the text differs substantially. The text for the experimental participants in baseline medical reads as follows: there are two players, a doctor (he) and a patient (she). The doctor chooses a price for the normal intravenous drip or the special intravenous drip. The patient is informed about the prices posted by the doctor and can decide if she wants to be examined or not. If not, this round ends. If they do, the doctor conducts a diagnosis in which he learns for certain the severity of the illness of the patient.[1] The doctor charges a price, specified in decision 1, from the patient for administering an intravenous drip. The price charged does not necessarily have to be the price for the intravenous drip actually chosen by the doctor in decision 3; it can also be the price of the other intravenous drip (for a complete set of instructions on the neutral and medical framing experiment see Appendix D.1).

3.3.1 Experimental Treatments

The three conditions in the neutral as well as in the medical framing constitute a $2 \times 2 \times 2$ factorial design, where the first '2' relates to the framing of the problem neutral or medical; the second and third '2' relate to whether *liability* and/or *verifiability* applies. Dulleck et al. (2011) and Kerschbamer et al. (2009) found in previous experiments the following values well suited for their analysis. In all treatments, this study let the customer's likelihood of requiring the major treatment be $h = 0.5$, and the value of sufficient treatment be $v = 10$. The costs of providing the minor, respectively major, treatments are $c_l = 2$, and $c_h = 6$. The prices posted by the sellers, p^l and p^h are integers from an interval $I \in \{1, 2, \ldots, 11\}$ with $p^l \leq p^h$. The outside option, if no trade takes place between the seller and the customer, is set to $o = 1.6$. Therefore, for this

[1] For simplicity we assume a diagnosis is free of charge.

project, the same strategy is followed.

The plan is to run all treatments with a total of 16 periods. Due to the repetition of the stage game, the matching of subjects is very important. It must not be possible for any seller to build up a reputation in the course of the repeated interaction (see Dulleck et al., 2011, for a credence goods experiment with reputation). This prohibits the use of a partner matching (in which a seller is matched with the same customer in all 16 periods) in all treatments; in fact a stranger matching is employed, in which customers and sellers are randomly rematched after each period. This will considerably reduce the ability to build up reputation but it cannot be completely ruled out since there is not a perfect stranger matching. In all treatments, eight subjects are in the role of customers (patients), and eight in the role of sellers (doctors). The assignment to roles is randomly determined at the beginning of the experiment, and roles are kept fixed throughout the entire experiment. Subjects are informed about the size of matching groups in each treatment. Hence subjects know that the probability of being matched with a specific trading partner in a specific round is one quarter. However, in all treatments, customers cannot identify sellers. Each matching group yields one statistically independent observation. The experiments were conducted in Austria at the University of Innsbruck in the winter semester 2012/2013 and October 2014.

3.3.2 Experimental Procedure

All experimental sessions were conducted using *CORAL* (Schaffner, 2013).[2] Recruiting was done with *ORSEE* in the winter semester 2012/2013 (Greiner, 2015) and *hroot* in October 2014 (Bock et al., 2012). A total of 240 ($= 15 \times 16$) students participated in the experiment conducted in 2012/2013 and 2014. Two sessions were run with neutral framing in condition $\mathbf{B}, \mathbf{L}$ and $\mathbf{V}$, since this is

[2]For a detailed introduction of CORAL or to have a look at the homepage that was particularly created for new users see Schaffner (2013) or `https://github.com/mas802/coral-econ` respectively.

only a replication of Dulleck et al. (2011), and three sessions with medical framing, which is the novelty of this study, in condition $\mathbf{B}, \mathbf{L}$ and $\mathbf{V}$. In all sessions we started with reading out the instructions to the participants. We emphasised in the instructions that all participants receive the same information[3]. The average session length was around 50 minutes. All participants in the experiment received an initial endowment of 10 points to start with and an exchange rate of 4 points equal to 1 euro was applied in the experiments. The average total payoff was 10.8 euro (sd = 4.22) in the medical framing and 10.1 euro (sd = 4.52) in the neutral framing environment.

3.3.3 Theoretical Predictions [4]

Standard economic theory assumes that both sellers and consumers are rational, risk neutral and own-money-maximising agents and that this is common knowledge. The equilibrium concept we apply is a perfect Bayesian equilibrium since we have a dynamic game with incomplete information. Our focus lies on symmetric equilibria.

PR1 (BN: *Baseline - Neutral Framing*): In markets without liability and/or verifiability (condition **BN** & **BM**), standard economic theory predicts that the expert seller would always provide the low treatment t^l and charge the customer for the high treatment t^h under each price vector. By knowing this, the customer would never enter the market if $p^h > 3$. However, with such a low p^h not even sellers that cheat (i.e undertreat) would make a profit ($p^h - c^l \leq 3 - 2 = 1 < 1.6 = o$) and therefore, we would expect to see a market collapse since there is no price such that a seller and a customer would interact and obtain at least as much as the outside option. Hence, we would assume to observe a market breakdown similar to the market for lemons described by Akerlof (1970). In other words *undertreatment* is a considerable inefficiency concern in the baseline condition. Additionally, *overtreatment* and *overcharging*

[3]See Appendix D.1.1 and D.1.2 for the experimental instructions for condition **BN** & **BM**.
[4]Taken from Dulleck et al. (2011) except for baseline medical framing.

are going to be an inefficiency concern in condition **BN** as well. Hence, the three inefficiency concerns in condition **BN** are *undertreatment*, *overtreatment* and *overcharging*.

PR2 (BM: *Baseline - Medical Framing*): Consider condition **BM** with ***Medical Framing***. Rational choice theory requires that preferences between possibilities do not shift when the framing of the experiment is altered. Therefore, theory would predict the same outcome in the medical framing as in the neutral framing discussed previously for condition **BN**. However, Charness (2010) states that participants are of great risk to the framing of experiments as they bring personal attitudes and beliefs to the laboratory, which they apply to situations they are unfamiliar with. Eckel and Gintis (2010) continue to state that other regarding preferences are driven by social norms, and social norms are environment specific. Expected utility theory states that participants should not exhibit preference reversal in the medical framing, but when taking into consideration these statements regarding social norms and prior experiences, we could expect that participants would behave more honestly in the medical framing as they can relate more to such a situation compared to the neutral framing in our condition **BN**.

PR3 (L: *Liability - Medical Framing and/or Neutral Framing*): In condition **L**, undertreatment is impossible as the expert seller is liable for the treatment he provides and *overtreatment* is dominated by *overcharging*. Therefore, an expert always provides the appropriate treatment and charges the price for the higher one under each price combination. Expecting such a behaviour, the consumer only interacts in the market if $p^h \leq 8$. With $p^h = 8$ and expected cost for providing the appropriate treatment equal to 4 ($= (1-h) \cdot c^l + h \cdot c^h$), the expected profit for the expert is 4, which is larger than the outside option of 1.6. Therefore, an expert always charges $p^h = 8$ while the price for the lower treatment is not determined (because it is never chosen by the expert since $p^l < p^h$). In other words, one would expect the market to generate high efficiency and maximum trade.

PR4 (V: *Verifiability - Medical Framing and/or Neutral Framing*): Due to the fact that a customer is able to verify ex-post the quality of the treatment she receives, an expert always charges the price of the treatment he provides (i.e. if he provides the low treatment he charges the price of the low treatment and conversely for the high treatment). Therefore, an expert provides the appropriate treatment under equal mark-up prices, which are defined such that $p^h - p^l = c^h - c^l = 4$. In condition **V**, the inefficiency concern of *overcharging* is therefore prevented but it does not necessarily eliminate *undertreatment* or *overtreatment*. Nonetheless, we would expect the market to generate high efficiency and maximum trade as the expert seller posts equal mark-up prices and always provides the appropriate treatment.

3.4 Results

3.4.1 Descriptive Summary Statistics

3.4.1.1 The Effect of Medical Framing on Baseline

Tables 3.4.1 and 3.4.2 present the descriptive results of all experimental treatments conducted in neutral and medical framing.

Table 3.4.1: Overview of Results - Neutral Framing

Averages per Period	**Baseline**	**Liability**	**Verifiability**
Trade[a]	0.48	0.67	0.46
Efficiency[b]	0.05	0.66	0.10
Undertreatment[a,c]	0.67	—	0.43
Overtreatment[a,d]	0.11	0.06	0.19
Overcharging[a,e]	0.83	0.72	—
p^l with trade	4.33	4.37	5.59
p^l without trade	4.90	5.75	6.32
p^h with trade	7.06	7.53	8.14
p^h without trade	8.16	9.26	8.94
Actually charged price	6.92	7.22	7.10
Profit sellers[f]	2.82	2.69	2.45
Profit consumers[f]	0.52	2.39	1.03
	(6,8) 17.6%	(6,8) 14.5%	(7,8) 10.6%
Most prominent price vector	(5,8) 15.6%	(5,8) 11.7%	(6,8) 9.8%
	(4,8) 6.6%	(4,7) 6.3%	(6,6) 7.8%
Number of obs. (# subjects)	512 (32)	512 (32)	512 (32)

[a] relative frequency.
[b] calculated as (actual average profit - outside option) / (maximum possible average profit - outside option)
[c] consumer needs t^h, but seller provides t^l.
[d] consumer needs t^l, but seller provides t^h.
[e] seller provides t^l, but charges p^h (with $p^l \leq p^h$ and consumer requiring t^l).
[f] in experimental currency units, in medical framing sellers = doctors and consumers = patients.

Table 3.4.2: Overview of Results - Medical Framing

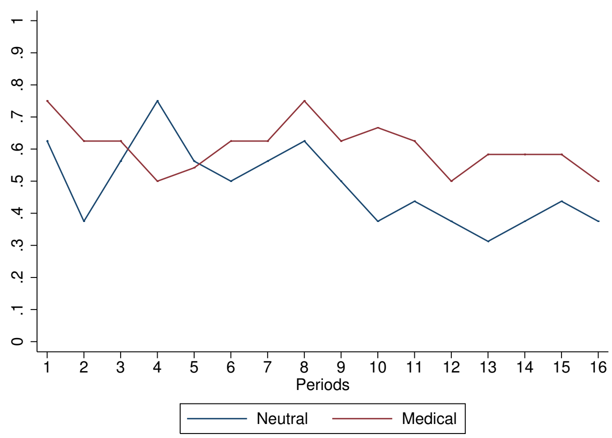
Averages per Period	**Baseline**	**Liability**	**Verifiability**
Trade	0.61	0.65	0.54
Efficiency	0.40	0.64	0.21
Undertreatment	0.30	—	0.47
Overtreatment	0.01	0.03	0.25
Overcharging	0.72	0.64	—
p^l with trade	4.62	4.50	5.54
p^l without trade	5.17	5.11	6.10
p^h with trade	7.83	7.85	8.21
p^h without trade	8.67	9.13	9.05
Actually charged price	7.43	7.31	6.76
Profit doctor	3.09	2.60	2.55
Profit patient	1.20	2.31	1.25
	(6,8) 18.0%	(6,8) 20.0%	(6,8)15.6%
Most prominent price vector	(4,8) 14.6%	(4,8) 20.0%	(7,8) 11.7%
	(5,8) 11.5%	(5,8) 11.5%	(5,9) 9.6%
Number of obs. (# subjects)	768 (48)	768 (48)	768 (48)

Note: See previous table's footnotes for a definition of the variables.

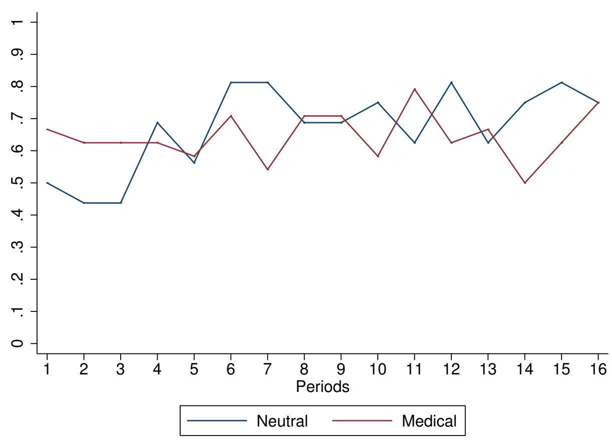
Economic theory would have predicted that in a market without liability and/or verifiability we would observe a market breakdown. However, this is not what we see. In fact we observe that, in *Neutral Framing*, there is a relative frequency of trade of 48% and in *Medical Framing* trade is even higher with 61%. Furthermore, the relative frequency of undertreatment of 67% indicates that in markets without a liability restriction, undertreatment is a considerable dilemma but not as pronounced as predicted by **PR1**. Considering all results we can therefore reject **PR1** since we do not observe a market breakdown or undertreatment to occur each time trade takes place. Nonetheless, a substantial number of times participants in the role of experts play according to standard economic theory by charging the high price p^h and delivering the low quality t^l. Additionally, the relative frequency of overcharging in condition **BN** is only partially in line with **PR1**. Subjects in the role of experts provided the low treatment and charged for the higher one in 83% of times, but not on every occasion as predicted by **PR1**.

Furthermore, **PR2** is rejected as well, since significant differences in the means between *Neutral framing* & *Medical framing* are observed when a Wilcoxon Rank-Sum Test is used to analyse the results as depicted in Table 3.4.3. For instance, all variables but one, p^l with trade, are significantly different from the neutral framing at either the 1% or 5% level of significance. Precisely, experimental participants in condition **BM** exhibit a more pro-social behaviour in the direction of being more honest. They undertreat, overtreat and overcharge significantly less ($p < 0.01$) compared to participants in condition **BN**. These significant differences are depicted in Figure 3.4.1 (a) for Trade, in Figure 3.4.3 (a) for Undertreatment, in Figure 3.4.4 (a) for Overtreatment and in Figure 3.4.6 (a) for Overcharging. Furthermore, participants allocated the roles of sellers ($p < 0.05$) and consumers ($p < 0.01$) in condition **BM** earn significantly more on average than participants in the equivalent roles in condition **BN**. The stronger pro-social behaviour of participants in the medical framing is further strengthened by their choice of price vectors. In the medical framing the equal mark-up price (4,8), is 8% more often selected than in the neutral framing.

Furthermore, when the equal mark-up price was chosen, in the medical framing environment, participants behaved honestly (i.e. no undertreatment) in more than 56% compared to only 11% in the neutral framing. However, in both conditions **BN** and **BM**, the participants allocated the role of consumers should not have entered the credence goods market as they earned (**BN** = 0.52 and **BM** = 1.20) considerably less than the outside option of 1.6.

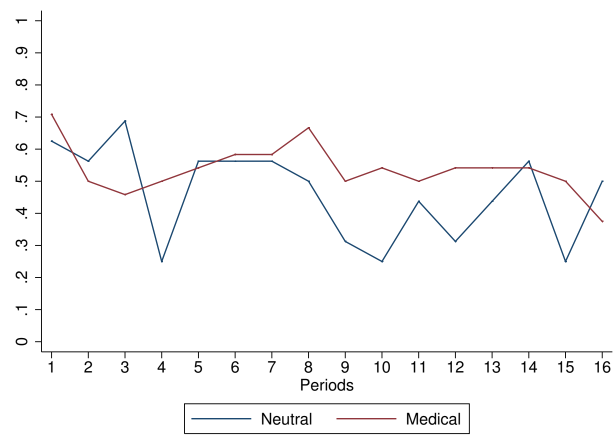


(a) Relative Frequency of Trade Baseline



(b) Relative Frequency of Trade Liability

Figure 3.4.1: Frequency of Trade Neutral vs Medical



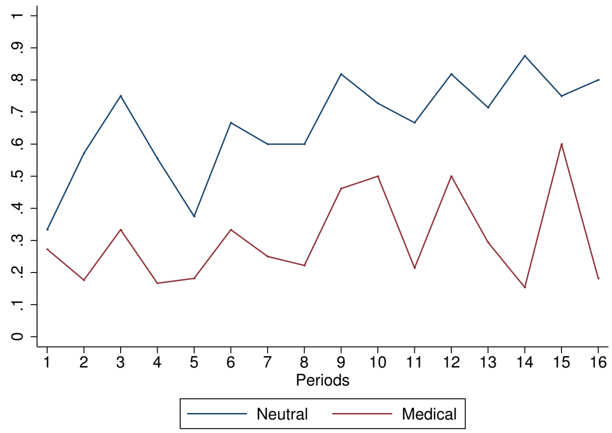
(a) Relative Frequency of Trade Verifiability

Figure 3.4.2: Frequency of Trade Neutral vs Medical

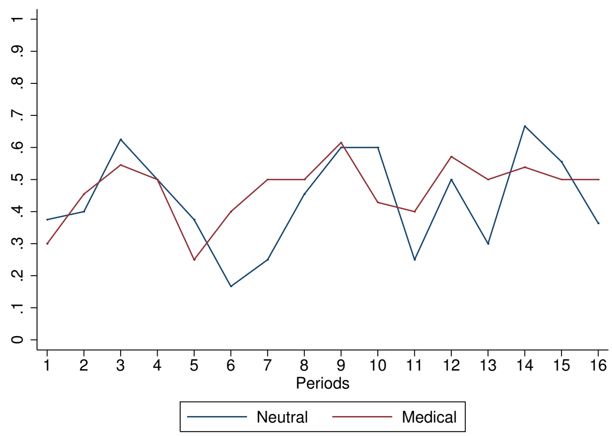
Table 3.4.3: Overview Ranksum Test Results - Medical vs Neutral Framing

	Medical - Neutral Framing, Mean Diff. ($\|z - value^{g}\|$)		
Averages per Period	Baseline	Liability	Verifiability
Trade[a]	0.13 (4.32)***	-0.02 (0.96)	0.08 (2.65)***
Efficiency[b]	0.34 (16.88)***	-0.02 (2.49)**	0.11 (6.38)***
Undertreatment[a,c]	-0.37 (9.39)***	—	0.04 (1.00)
Overtreatment[a,d]	-0.10 (5.43)***	-0.03 (1.63)	0.06 (1.70)*
Overcharging[a,e]	-0.11 (3.98)***	-0.08 (2.04)**	—
p^l with trade	0.29 (1.58)	0.13 (0.59)	-0.05 (1.19)
p^l without trade	0.27 (2.24)**	-0.64 (3.28)***	-0.22 (1.98)**
p^h with trade	0.77 (7.90)***	0.32 (4.09)***	0.07 (0.18)
p^h without trade	0.51 (5.01)***	-0.13 (1.85)*	0.11 (0.32)
Actually charged price	0.51 (4.76)***	0.09 (1.90)*	-0.34 (2.49)**
Profit sellers[f]	0.27 (2.35)**	-0.09 (1.68)*	0.10 (1.73)*
Profit consumers[f]	0.68 (12.35)***	-0.08 (3.86)***	0.22 (4.08)***

*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.
[a] relative frequency.
[b] calculated as (actual average profit - outside option) / (maximum possible average profit - outside option)
[c] consumer needs t^h, but seller provides t^l.
[d] consumer needs t^l, but seller provides t^h.
[e] seller provides t^l, but charges p^h (with $p^l \leq p^h$ and consumer requiring t^l).
[f] in experimental currency units, in medical framing sellers = doctors and consumers = patients.
[g] Wilcoxon Rank-Sum Test where Ho: variable$_i$(medical==0) = variable$_i$(medical==1).



(a) Undertreatment Baseline

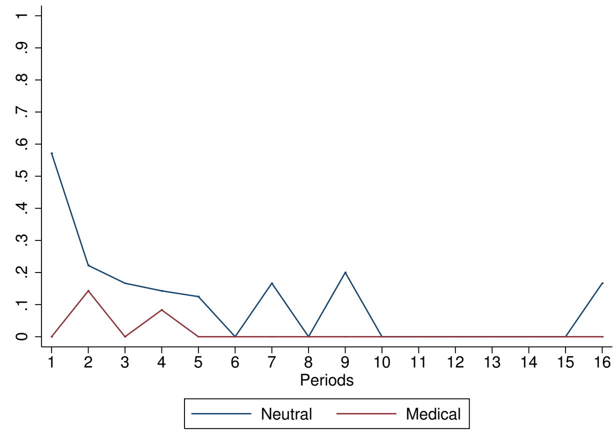


(b) Undertreatment Verifiability

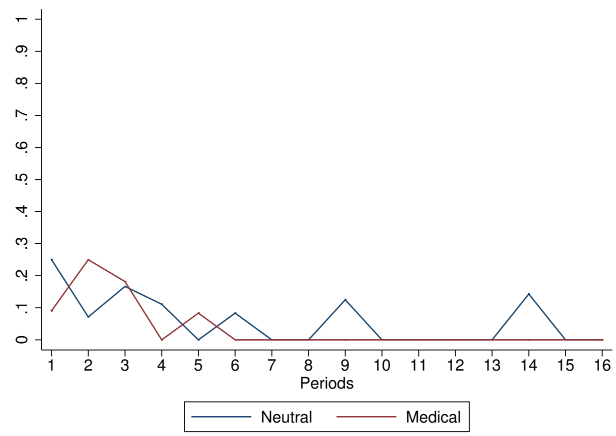
Figure 3.4.3: Average Rate of Undertreatment Neutral vs Medical

3.4.1.2 The Effect of Medical Framing on Liability

As participants in the liability condition always receive a sufficient treatment, economic theory predicts that they should always participate in the market. In particular, in condition **L** undertreatment is impossible, hence, we should observe high efficiency and maximum trade to occur as described in **PR3**. In line with **PR3** we observe a sizeable increase in trade for both *neutral* and *medical*



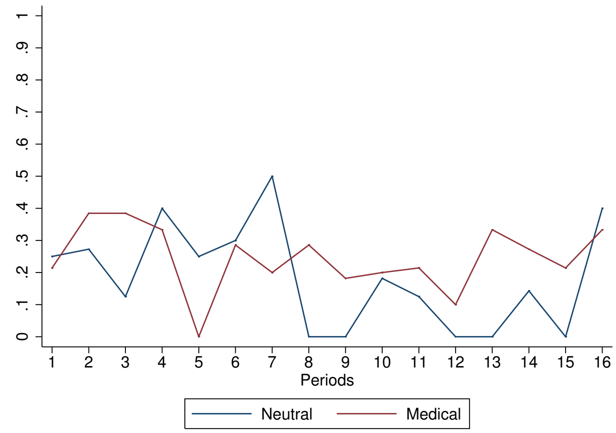
(a) Overtreatment Baseline



(b) Overtreatment Liability

Figure 3.4.4: Average Rate of Overtreatment Neutral vs Medical

framing compared to the baseline treatment, however, we do not observe maximum trade as predicted in **PR3**. Moreover, we also see a considerable increase in efficiency as expected compared to the baseline condition. On the other hand, the relative frequency of overcharging in both neutral and medical framing of 72% and 64% respectively is below the predicted 100% in **PR3**. In fact it even decreased compared to the baseline condition. Taking into consideration all

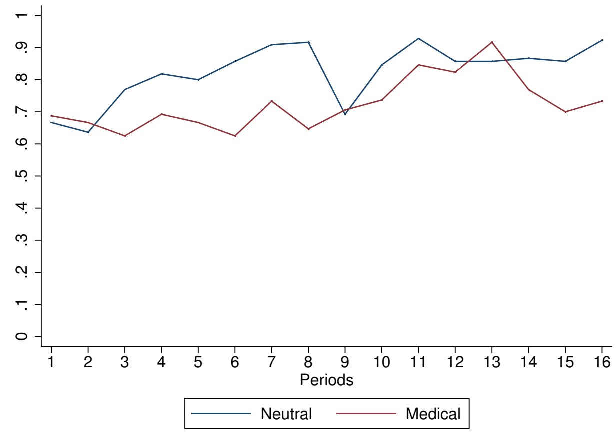


(a) Overtreatment Verifiability

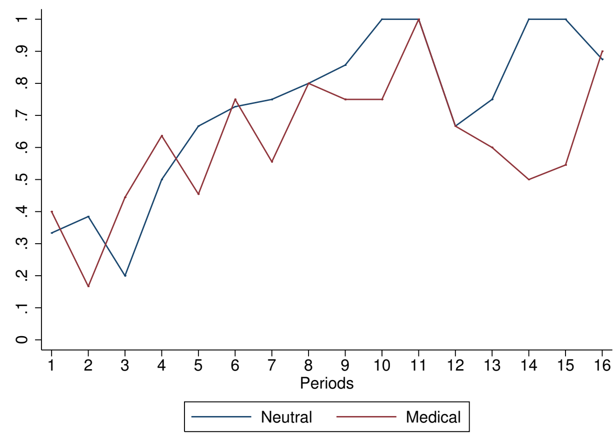
Figure 3.4.5: Average Rate of Overtreatment Neutral vs Medical

results, we observe that participants in the liability treatment move towards a behaviour that is more consistent with the predictions in **PR3**.

As previously mentioned in **PR2** rational choice theory requires that participants in a laboratory experiment do not shift their preferences when a different framing to the same experiment is applied. When we compare the neutral with the medical framing in the liability condition, we do not observe such a strong effect of medical framing on the behaviour of the participants as in the baseline condition. Nonetheless, we still observe significant differences between the neutral and medical framing for variables such as overcharging, efficiency and consumer's profit which is in contradiction to **PR2** (see Table 3.4.3 for a complete overview). On the other hand, in line with **PR2**, for variables such as trade and overtreatment we do not find any significant differences. This can be seen in Figure 3.4.1 (b) and 3.4.4 (b) for trade and overtreatment respectively, or the insignificant Wilcoxon Rank-Sum test statistic for trade and overtreatment in Table 3.4.3 (middle column). Additionally, in contrast to the baseline condition, the introduction of the liability restriction achieved a substantial improvement in regards to the payoff for consumers. Not only do the sellers (2.69, 2.60) earn more than the outside option of 1.6 as in the baseline condition, but also the



(a) Overcharging Baseline



(b) Overcharging Liability

Figure 3.4.6: Average Rate of Overcharging Neutral vs Medical

consumers (2.39, 2.31) earn more in both framings. One noticeable result with the introduction of medical framing in liability, is the fact that participants in the role of sellers chose the equal mark-up price (4,8) in 20% of periods. This is an increase of almost 16%[5] compared to the liability condition with neutral framing.

[5]The price vector (4,8) in liability with neutral framing is chosen in 4.30% of observations (not shown in the table).

3.4.1.3 The Effect of Medical Framing on Verifiability

Including the verifiability restriction in a credence goods market will solve the problem of overcharging but it does not prevent the market inefficiencies of undertreatment and overtreatment. In particular, overcharging is not possible because the consumer can observe and verify ex-post the quality provided by the expert. Hence, the expert seller is unable to charge for the high treatment c^h if he has provided the low treatment c^l. Standard economic theory predicts a market that generates maximal trade volume and full efficiency. Contrary to **PR4** in neutral and medical framing we do not observe maximum trade or high efficiency to occur. Moreover, the participation rate in the market where verifiability is introduced is even lower compared to our baseline condition with no institutional restrictions at all (see Table 3.4.1 for neutral framing and Table 3.4.2 for medical framing respectively).[6] The equal mark-up vectors predicted in **PR4** only happen to occur in around 5% of cases. Similar to the liability condition we do not find a very strong effect, with some exceptions, in verifiability when we alter the context from an abstract to a medical framing. We do find, however, a significant increase at the 1% level of trade and efficiency in medical framing as depicted in Table 3.4.3. In addition, consumers in medical framing earn significantly more ($p < 0.01$) than consumers in neutral framing. Nonetheless, consumers in both environments of the experiment should not have participated in the market as the average earnings are below the outside option of 1.6. Hence, it seems that in a credence goods experiment, contrary to the theoretical predictions by Dulleck and Kerschbamer (2006), verifiability has at most a minor effect on the behaviour of participants, which is line with a previous study by Dulleck et al. (2011) who also find a similar result.

To obtain a clearer picture of how the behaviour of the participants evolved over the duration of the experiment, the findings of the different environments over time are graphically illustrated. Examining first the relative frequency of

[6] For a comparison between the baseline and the verifiability condition see Table A.2.1 in Appendix A.2.

trade, we can depict that in baseline (Figure 3.4.1 (a)) and verifiability (Figure 3.4.2), compared to liability (Figure 3.4.1 (b)), the participation rate in the market decreases over time. One reason for such a steady decline in the frequency of trade is the fact that participants in the role of consumers learn that undertreatment occurs and that it is costly for them. Furthermore, as depicted in Figure 3.4.1 (a) medical framing has a positive effect on the overall participation rate of subjects' propensity to interact in the market. In conjunction with the observable decrease in trade in the neutral framing environment, we also see in Figure 3.4.3 (a) a steady increase in undertreatment over time. In particular, in the baseline neutral framing environment we observe an increasing trend from 30% to 80% in the propensity to undertreat over time, while in medical framing undertreatment fluctuates around 30%. This is confirming the significant ($p < 0.01$) Wilcoxon Rank-Sum test result in Table 3.4.3 for the difference between neutral and medical framing for undertreatment in the baseline condition. The behaviour over time for the baseline neutral case is also indicating that participants in the experiment move closer towards an outcome predicted by economic theory, in particular, less participation in the market and more undertreatment. When we examine Figure 3.4.6 (a) it can be seen that the frequency to overcharge in the baseline condition starts in period one at around 70% for both framings. However, in the following periods, the behaviour of the two participant groups differs. While in neutral framing the frequency of overcharging is increasing to almost 90% over the 16 periods, for medical framing it increases to almost 90% after 13 periods but then drops again to 70%. We also observe a significant difference ($p < 0.01$) between the two environments when we test for it with a Wilcoxon Rank-Sum test (see Table 3.4.3). Examining Figure 3.4.6 (b) we observe that participants' tendency to overcharge in the liability condition, occurs in around 40% of cases in period one for both framings, after which, it increases with some fluctuations to almost 100% in the last period. This is perfectly in line with economic theory that expert sellers always overcharge in the liability condition. The frequency of overtreatment presents a mixed result. While we observe a significant difference in the means for the neutral and medical framing in the baseline condition presented in Table 3.4.3, we can also depict in Figure 3.4.4 that this result is driven by the first few rounds of the experiments, after which the participants in the two different environments behave alike. For liability we cannot really observe a different behaviour of the participants between the two framings, which is supported by the insignificant Rank-Sum test result in Table 3.4.3. The behaviour of participants in medical framing is usually more pro-social except for verifiability, in which we find a positive, weakly significant difference in the tendency to overtreat, meaning that participants in medical framing overtreat more, compared to participants in neutral framing.[7]

3.4.2 Estimating the Effect of Medical Framing on Trade, Undertreatment, Overtreatment and Overcharging

As our dependent variables - trade, undertreatment, overtreatment and overcharging - are binary, we utilise a probit model for our multivariate analysis. The outcome variable is determined by whether the observation came from a session with medical framing (=1) or neutral framing (=0). Additionally, we include a variable each for the asking price of the low and high treatment respectively. To account for the learning effect during the experiment we divided the period variable into three different segments with Period 9-16 as our reference group. We also calculated the marginal effects to interpret the magnitude of the effects.

Table 3.4.4 presents the results from a probit estimation where we analyse the effect of medical framing, the price of the respective treatment and the learning effect over the 16 periods on the relative frequency of trade for our baseline, liability as well as verifiability restrictions. Examining our main variable of interest, medical framing, we see that it has a very strong positive

[7] To observe the same analysis for undertreatment, overtreatment and overcharging with individual participant averages, please refer to Appendix A.1. The results for the baseline condition are still highly significant, while for liability or verifiability the mean differences are insignificant.

influence on the frequency of trade. Participants in medical framing are 23.8 percentage points more likely to participate in the market, contrary to liability or verifiability where medical framing has no significant effect at all. Moreover, we observe that the asking price for the low treatment has a significant effect ($p < 0.01$) on the relative frequency of trade in liability and verifiability but is insignificant in our baseline condition. The price for the high treatment has an even stronger effect on the frequency to interact in the market. For all conditions baseline, liability and verifiability, the asking price for the high treatment is significant at the 1% level. The marginal effect is also quite strong. For instance, when we inspect the baseline specification, for every lab$ increase in the high price, participants are 18.6 percentage points less likely to interact in the market. Lastly, when we examine how the propensity to interact changes over time, we observe that for the baseline and verifiability condition there is a significant positive difference ($p < 0.01$) in the first eight periods (Period 1-3 and Period 4-8) to participate in the market, compared to our reference group Period 9-16. For liability, we observe a reverse pattern. In Periods 1-3 there is lower participation rate than for the later periods, indicating that participants realise that they always obtain a sufficient treatment since experts are not able to undertreat a consumer.

Moving on to Table 3.4.5 we examine the impact of our independent variables, Medical Framing, Price Low, Price High and the different period ranges Period 1-3, Period 4-8 and Period 9-16, our reference group, on the relative frequency of undertreatment. Our main variable of interest, medical framing, is highly significant ($p < 0.01$) in altering participants propensity to undertreat in the baseline treatment. In medical framing, undertreatment is 44.5 percentage points less likely to occur compared to our reference group neutral framing. This is in clear contrast to our verifiability condition, where we cannot find any significant effect of medical framing on the likelihood to undertreat. Besides, the behaviour of participants in the verifiability condition does not change over the 16 periods. There is no significant effect between the early periods, Periods 1-3 and Periods 4-8, and our reference group Period 9-16.

We next focus our analysis on the impact of medical framing and the other independent variables on the relative frequency of overtreatment (see Table 3.4.6). For medical framing we see a similar result as for undertreatment. Only in the baseline treatment do we detect a significant impact ($p < 0.01$) of medical framing on participants' tendency to overtreat. Specifically, in medical framing, overtreatment is 6.6 percentage points less likely to occur compared to the reference group neutral framing. Additionally, in baseline and liability, one can also see that the likelihood of overtreatment is significantly higher in earlier periods than in the reference group Period 9-16. For verifiability we do not find any significant effect for medical framing or a changing behaviour of subjects in later periods of the experiment.

In the last part of this section we focus our investigation on Table 3.4.7 where we present probit regression results for the effect of medical framing on overcharging in our baseline and liability conditions. The negative coefficient for medical framing is highly significant at the 1% level in both conditions. For instance, in our baseline condition, overcharging is 14.7 percentage points less likely to occur in the medical environment compared to the neutral environment. A comparable effect of medical framing on the relative frequency of overcharging is observed in liability, where participants are 15.8 percentage points less likely to provide the low treatment and charge for the expensive one. Furthermore, examining the effect of Price low and Price high on the tendency to overcharge, one can see that in baseline the price of the low treatment has a significant negative influence on the overcharging behaviour of experts, while the price for the expensive treatment has a significant positive effect. This means that expert sellers in the experiments are 6.7 percentage points more likely to overcharge for every lab$ increase in the high price. A similar result can be observed for the liability restriction, such that only the quantitative effects differ slightly but not the level of significance. Lastly, investigating the propensity of overcharging uncovers that participants in the baseline condition behave significantly differently in the first three periods compared to later periods. A different picture is presented when we examine liability. The propensity to overcharge is significantly lower in Periods 1-3 and Periods 4-8 compared to our reference group 9-16.

Table 3.4.4: Probit Regression - Trade

Variables	Trade Baseline	Trade Liability	Trade Verifiability	Trade Overall
Medical	0.609***	-0.063	0.205	0.259***
	(4.24)	(-0.45)	(1.47)	(3.20)
	0.238	*-0.022*	*0.082*	*0.101*
Price Low	-0.043	-0.113***	-0.156***	-0.102***
	(-1.39)	(-3.74)	(-4.90)	(-5.88)
	-0.017	*-0.039*	*-0.062*	*-0.040*
Price High	-0.473***	-0.643***	-0.372***	-0.433***
	(-11.69)	(-9.41)	(-7.36)	(-14.76)
	-0.186	*-0.223*	*-0.148*	*-0.169*
Period 1-3	0.585***	-0.635***	0.596***	0.252***
	(5.95)	(-5.29)	(5.43)	(3.54)
	0.217	*-0.236*	*0.230*	*0.096*
Period 4-8	0.480***	-0.031	0.452***	0.325***
	(6.27)	(-0.31)	(4.68)	(6.25)
	0.184	*-0.011*	*0.178*	*0.124*
Liability				0.439***
				(4.39)
				0.167
Verifiability				0.230**
				(2.30)
				0.089
Constant	3.507***	6.523***	3.737***	3.789***
	(10.74)	(10.15)	(6.93)	(14.32)
χ^2	198.095***	165.605***	94.723***	327.197***
$Prob > \chi^2$	0.000	0.000	0.000	0.000
(Pseudo) R^2	0.146	0.257	0.126	0.153
N (Clusters)	80	80	80	240
N (Obs.)	1280	1280	1280	3840

Notes: Clustered standard errors, z-statistics in parentheses and marginal effects in italics. The reference groups consist of NEUTRAL, PERIOD 9-16 and BASELINE. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 3.4.5: Probit Regression - Undertreatment

Variables	Undertreatment Baseline	Undertreatment Verifiability	Undertreatment Overall
Medical	-1.183***	0.309	-0.351**
	(-6.24)	(1.45)	(-2.51)
	-0.445	*0.120*	*-0.139*
Price Low	-0.441***	0.433***	0.011
	(-7.13)	(6.32)	(0.26)
	-0.174	*0.170*	*0.004*
Price High	0.149***	-0.487***	-0.196***
	(2.60)	(-6.84)	(-4.36)
	0.059	*-0.191*	*-0.078*
Period 1-3	-0.574***	0.125	-0.213**
	(-3.58)	(0.88)	(-2.11)
	-0.213	*0.050*	*-0.083*
Period 4-8	-0.527***	-0.066	-0.236***
	(-4.06)	(-0.57)	(-3.01)
	-0.201	*-0.026*	*-0.092*
Verifiability			0.124
			(0.88)
			0.049
Constant	1.750***	1.251**	1.687***
	(4.22)	(2.27)	(5.56)
χ^2	117.732***	61.385***	60.697***
$Prob > \chi^2$	0.000	0.000	0.000
(Pseudo) R^2	0.257	0.211	0.056
N (Clusters)	80	80	160
N (Obs.)	666	648	1314

Notes: Clustered standard errors, z-statistics in parentheses and marginal effects in italics. The reference groups consist of NEUTRAL, PERIOD 9-16 and BASELINE. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 3.4.6: Probit Regression - Overtreatment

Variables	Overtreatment Baseline	Overtreatment Liability	Overtreatment Verifiability	Overtreatment Overall
Medical	-1.101***	-0.266	0.252	-0.191
	(-3.64)	(-1.20)	(1.45)	(-1.62)
	-0.066	*-0.015*	*0.063*	*-0.026*
Price Low	0.178**	-0.117	-0.407***	-0.192***
	(2.37)	(-1.63)	(-6.14)	(-5.37)
	0.007	*-0.006*	*-0.105*	*-0.026*
Price High	-0.063	-0.066	0.388***	0.129***
	(-0.74)	(-0.73)	(6.51)	(2.86)
	-0.003	*-0.004*	*0.100*	*0.017*
Period 1-3	1.091***	1.287***	0.166	0.785***
	(4.35)	(5.93)	(0.96)	(6.87)
	0.095	*0.161*	*0.045*	*0.146*
Period 4-8	0.494**	0.370*	0.187	0.299***
	(2.23)	(1.75)	(1.37)	(3.22)
	0.025	*0.024*	*0.050*	*0.043*
Liability				-0.107
				(-0.59)
				-0.014
Verifiability				1.072***
				(6.46)
				0.188
Constant	-2.171***	-1.094	-2.147***	-2.034***
	(-4.32)	(-1.62)	(-3.81)	(-5.61)
χ^2	60.297***	116.805***	76.923***	193.817***
$Prob > \chi^2$	0.000	0.000	0.000	0.000
(Pseudo) R^2	0.259	0.190	0.227	0.192
N (Clusters)	80	80	80	240
N (Obs.)	614	624	632	1870

Notes: Clustered standard errors, z-statistics in parentheses and marginal effects in italics. The reference groups consist of NEUTRAL, PERIOD 9-16 and BASELINE. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 3.4.7: Probit Regression - Overcharging

Variables	Overcharging Baseline	Overcharging Liability	Overcharging Overall
Medical	-0.717***	-0.479***	-0.582***
	(-3.91)	(-2.61)	(-4.45)
	-0.147	*-0.158*	*-0.157*
Price Low	-0.696***	-0.295***	-0.473***
	(-8.48)	(-5.37)	(-9.04)
	-0.148	*-0.100*	*-0.132*
Price High	0.327***	0.147***	0.251***
	(4.14)	(2.82)	(5.10)
	0.069	*0.050*	*0.070*
Period 1-3	-0.450**	-1.193***	-0.789***
	(-2.38)	(-6.86)	(-6.46)
	-0.112	*-0.444*	*-0.262*
Period 4-8	-0.145	-0.314**	-0.223**
	(-1.09)	(-2.54)	(-2.42)
	-0.032	*-0.110*	*-0.064*
Liability			-0.238*
			(-1.80)
			-0.068
Constant	2.167***	1.349***	1.629***
	(4.09)	(4.22)	(5.15)
χ^2	82.025***	105.089***	160.635***
$Prob > \chi^2$	0.000	0.000	0.000
(Pseudo) R^2	0.301	0.174	0.222
N (Clusters)	80	80	160
N (Obs.)	882	596	1478

Notes: Clustered standard errors, z-statistics in parentheses and marginal effects in italics. The reference groups consist of NEUTRAL, PERIOD 9-16 and BASELINE. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

3.5 Concluding Remarks

In this chapter, this study has examined how medical framing influences the behaviour of participants in a credence goods laboratory experiment on market specific variables such as trade, undertreatment, overtreatment and overcharging. The most striking result of the implementation of medical framing is that it significantly increases participants pro-social behaviour in our baseline condition. Specifically, undertreatment is almost 45% less likely to occur in medical framing compared to the neutral framing environment. However, when an institutional restriction is included, such as liability, which in itself is already a strong restriction of the players action space, i.e. no undertreatment possible, no strong differences are found between neutral and medical framing. In some cases, even a negative effect of medical framing on the behaviour of the participants is found. The same is true for verifiability, where the effect of medical framing is less prevalent or not significant at all. Why medical framing has no effect on verifiability is at this point unclear. For undertreatment, the difference between medical framing and neutral framing in the verifiability condition is insignificant, which is in strong contrast to the almost 45 percentage point decrease in the propensity to undertreat in the baseline condition. Hence, to obtain a clearer result, more experiments with naturalistic framings need to be undertaken in a credence goods setting.

Chapter 4

Medical Professionals as Participants in Credence Goods Experiments

4.1 Introduction

Supplier-induced demand (SID) is seen as a major issue in the provision of health services and has been well documented in the empirical literature (Fuchs, 1978; Gruber and Owings, 1996; Gruber et al., 1999; Jürges, 2007). All these studies rely on field data, with its limitation with respect to allowing a full identification of causal effects. The theory best suited to model the role of incentives in this case - Credence Goods theory (Darby and Karni (1973) or Dulleck and Kerschbamer (2006, for a comprehensive survey)) - has been tested in experiments (Kerschbamer et al., 2009; Dulleck et al., 2011; Beck et al., 2014) and by and large can confirm the empirical literature on SID. Kerschbamer et al. (2015) also raises the point that participants in these experiments are consistently driven by different motivations; some simply maximise own payoffs, others are driven by efficiency concerns or inequality aversion. In this article we use a credence goods experiment with medical professionals as participants to address one important question; do medical professionals behave better or worse compared to standard experimental participants, i.e does the selection into the health profession increase or mitigate the credence goods problems.

In health economics, SID refers to 'services ordered by a physician for a patient, that the patient would refuse if he or she had the same medical knowledge and expertise as the physician' (Hay and Leahy, 1982). SID is one of the reasons stated that total expenditure on health in total, but also as a percentage of gross domestic product, has increased for most countries around the globe. For example, between 1995 and 2012 for countries such as Germany, Austria, Australia or the US between 12% and 32% respectively.[1] Arrow (1963) argues that SID in healthcare markets results from the information asymmetry between a doctor and a patient. Bickerdyke et al. (2002) suggest two different reasons for the occurrence of SID, namely physicians' self-interest and the well-being of their patients. Furthermore, a study cited by Emons (1997) points to the fact that an average person has a one third higher likelihood of receiving one

[1] `http://apps.who.int/gho/data/node.main.75`, accessed 22.03.2015.

of seven major surgical interventions, compared to a doctor or a member of a doctor's family. Such evidence suggests that better informed and educated patients are less likely to be offered or to accept unnecessary and expensive surgical interventions. Fuchs (1978) shows that there is substantial evidence for the hypothesis that surgeons shift the demand for operations. In particular, if the surgeon population ratio increases, the demand for operations increases as well. Moreover, Jürges (2007) discovers indirect support for the hypothesis that private health insurance patients in Germany are convinced to more treatments by their physicians compared to statutorily insured patients, due to the higher fees chargeable for the former ones. Further evidence is provided by Gruber et al. (1999) who show that fee differences of health insurance programs between caesarean deliveries and normal child births, such that caesarean deliveries are more profitable in one health program, have an influence on the relative frequency of caesarean deliveries. Gruber and Owings (1996) also report that the frequency of caesarean section deliveries is negatively correlated with higher birth rates. Hughes and Yule (1992) find that the number of cervical cytology treatments is positively correlated with the fee for this treatment. Furthermore, the fear of doctors being held liable for malpractice could also increase the tendency of physicians to overtreat (Kessler and McClellan, 1996).[2]

In information economics, markets of asymmetric information between an expert and a consumer, such as a doctor and a patient, have been classified as credence goods markets (Darby and Karni, 1973). The term credence good describes a situation in which an expert knows more about the quality of a good a consumer needs than the consumer herself (Dulleck and Kerschbamer, 2009).[3] Darby and Karni (1973) added this type of good to Nelson's (1970) classification of ordinary, search and experience goods.

In a medical consultation, the patient knows that something is wrong but she cannot determine the exact gravity of her illness. Even though the physician

[2]While costly from a social perspective overprescribing healthcare services does not necessarily have to be a disadvantage for the patient: Folland et al. (2013) states that 'inducing more care does no harm if it encourages a move toward the patient optimum' (p. 306).

[3]For a survey of the credence goods literature see Dulleck and Kerschbamer (2006).

will make a diagnosis and suggest a treatment, the patient cannot verify if the prescribed treatment is the appropriate one given the diagnosis. Such information asymmetries between a doctor and a patient can lead to inefficiencies such as undertreatment, overtreatment and overcharging. Furthermore, a key feature of credence goods is that a patient, even ex-post, cannot verify with certainty if she received the appropriate service. Hence, the impossibility of verifying the exact nature of the patients problem gives dishonest doctors the opportunity to mistreat.

Despite such influence of doctors, not many studies exist that have investigated physicians' incentives to provide the appropriate treatment in a laboratory environment. The utilisation of laboratory experiments in health economics has only recently emerged, most of which have analysed payment systems and how they affect the incentives of the participants in the experiments by employing non-medical and medical students. (e.g. Hennig-Schmidt et al., 2011; Kairies and Krieger, 2013; Brosig-Koch et al., 2013; Green, 2014; Hennig-Schmidt and Wiesen, 2014; Keser et al., 2014). One exception is Brosig-Koch et al. (2014) who not only use students but also real physicians in an artefactual field experiment to analyse the behaviour of physicians between a fee-for-service and capitation system.[4] They find that the likelihood of physicians to overtreat is more widespread in the fee-for-service system. Another related paper is Kesternich et al. (2014) who use 132 medical students to study the effect of professional norms and how they influence the behaviour of prospective medical professionals in a laboratory experiment. They find that medical students, when they have been shown the Hippocratic Oath prior to playing the game, significantly increase their willingness to forgo some of their own monetary payoff to the advantage of another participant and also decrease their efficiency concerns in the experiment. However, the authors also state that professional norms may decrease the problem of asymmetric information but at the same time they can lead to more overtreatment. Furthermore, Huck et al. (2014) analysed in an

[4] "An artefactual field experiment is the same as a conventional lab experiment but with a nonstandard subject pool"(Harrison and List, 2004, p. 1014).

experimental setting how a medical insurance and the free choice of a physician correlates with the overtreatment of patients. They find that competition mitigates the problem of overtreatment, while having insurance encourages physicians' propensity to overtreat patients.

It is often argued that experiments with only a student participant pool lack external validity. Thus, investigating supplier provision behaviour in an abstract artefactual field experiment will provide us with important clues in regards to traits of physician's activities as suppliers in the health provision. Al-Ubaydli and List (2013) point out that 'the best predictor of cheating rates in the natural field experiment is behavior in sterile laboratory treatments with neutral language instructions' (p. 33). Hence, this study investigates the behaviour of real physicians in an abstract credence goods experiment and compares their behaviour to a standard participant pool. The results obtained with students from Chapter 3 are used as a reference point to which we compare our results for medical professionals. To the best of our knowledge, this research is the first that attempts to analyse the SID problem using a credence goods experiment with real physicians. We find that medical professionals undertreat and overcharge significantly less than students, however, at the same time, physicians overtreat significantly more than students.[5] The more frequent occurrence of real world experts' propensity to overtreat has previously been reported by Beck et al. (2014), by comparing the behaviour of car mechanics and students in a credence goods laboratory experiment as well as two recently published field experiments that analyse the credence goods market with real experts such as taxi drivers and car mechanics respectively (Balafoutas et al., 2013; Schneider, 2012). The remaining paper is structured as follows: the next section discusses the methodology, followed by the results and concluding remarks.

[5] To minimise a possible experimenter demand effect in experiments with medical professionals an external researcher, not familiar with the experiment, was handling most of the interactions and recruiting of participants.

4.2 Methodology

4.2.1 Model

4.2.1.1 Basic Setup

We start by utilising a simplified credence goods model developed by Dulleck and Kerschbamer (2006). In this game, there are two players: a doctor and a patient. In the first decision the doctor posts two prices, a price p^h for the high quality service and a price p^l for the low quality service, with the condition (that) $p^h \geq p^l$. The patient gets to know the prices for p^h and p^l and decides whether she wants to participate in the market or not[6]. If not, no trade takes place and both participants receive the outside option $o > 0$. If she decides to interact in the market, nature determines with probability h that the patient needs the high quality treatment and with probability $1 - h$ the low quality one. The doctor learns with certainty which treatment will solve the needs of the patient and either supplies the high quality treatment t^h or the low quality treatment t^l, where t^h incurs costs of c^h and t^l of c^l respectively, with $c^h \geq c^l$. In the final decision, the doctor charges one of the prices previously specified for the corresponding treatment. The patient earns a payoff $v > 0$ if the treatment chosen by the doctor was sufficient (*i.e.* if she needs the high quality and obtains the high quality, or she requires the low quality and either receives the high quality or the low quality) or *zero* otherwise. The patient, irrespective of receiving a sufficient or insufficient treatment, has to disburse the cost of the treatment administered. More formally, let $\theta \in \{l, h\}$ be the index of a consumers type of problem, $\delta \in \{l, h\}$ be the index for the quality of the treatment provided and let $\gamma \in \{l, h\}$ be the index of the quality of treatment actually charged for. The material payoffs of the seller π_d and consumer π_p are therefore as follows:

[6] We ignore the fact that a patient in a real live situation who is seriously ill cannot just walk away from visiting a doctor. In such circumstances the outside option should probably have a negative value as the patient is still not cured from her illness.

$$\pi_d(p^l, p^h, \delta, \gamma) \quad = \quad p^\gamma - c^\delta \tag{4.1}$$

$$\pi_p(p^l, p^h, \theta, \delta, \gamma) = \begin{cases} -p^\gamma & \text{if} \quad \theta = h \quad \text{and} \quad \delta = l, \\ v^\theta - p^\gamma & \text{otherwise.} \end{cases} \tag{4.2}$$

4.3 Experimental Design

The design of the experiment is the same as previously applied by Dulleck et al. (2011) and in a recently published paper by Beck et al. (2014). In this paper, only the baseline condition without institutional restriction, such as liability[7] or verifiability[8], specified as **B/N** in Dulleck et al. (2011) is used. The main differences in this paper are the utilisation of medical professionals, in particular General Practitioners (GP's) as participants as well as the implementation of a strategy method for the players allocated the role of doctors.

4.3.1 Experimental Treatment

Dulleck et al. (2011) and Kerschbamer et al. (2009) found in previous experiments the following parameters calibrated for their analysis; to be able to compare the results of this research with theirs, the same strategy is followed in this paper. In all treatments we let the patient's probability of needing the major treatment be h = 0.5, and the value of sufficient treatment be v = 10. The costs of providing the minor, respectively major, treatments are c_l = 2, and c_h = 6. The prices posted by the doctors, p^l and p^h are integers from an interval

[7] In liability, undertreatment is impossible as the expert seller is liable for the treatment he provides.

[8] Verifiability prevents the problem of overcharging because consumers can observe and verify ex-post the quality provided by the seller.

$I \in \{1, 2, \ldots, 11\}$ with $p^l \leq p^h$. The outside option if no trade takes place between the doctor and the patient is determined at $o = 1.6$ (see below). A total of 16 periods is run per session.

To avoid a doctor to building up reputation the matching of participants is very important. A stranger matching is employed, in which a doctor and a patient are rematched after each period. Compared to a partner matching, in which a doctor and a patient interact with each other for the whole 16 periods, this should considerably reduce the ability to build up reputation. It cannot completely be dismissed, however, since there is not a perfect stranger matching.

In all sessions conducted with students, matching groups of 8 participants each are employed, of which four are in the role of sellers, and four in the role of consumers.[9] The assignment to roles is randomly determined at the beginning of the experiment, and roles are kept fixed throughout the entire experiment. Participants are informed about the size of the matching groups in each session. Hence, participants know that the probability of being matched with a specific trading partner in a specific round is one quarter. However, in all sessions, participants in the role of experts cannot identify participants in the role of consumers. A strategy method was applied for the experiment as previously mentioned. The strategy method allows researchers to obtain a richer dataset as players in the experiment have to outline their strategy profile for a possible decision of the opposite player. This has substantial advantages since one can elicit preferences even at normally unattained decision nodes (Zizzo, 2010). Specifically, for this experiment it means that for every round an observation on the decision made of the expert seller was obtained, even for the cases when the consumer decided not to participate in the market. Figure 4.3.1 illustrates the sequence of decisions in this game and payoffs provided to each player.[10]

[9]For the experiments conducted with medical professionals we did not particularly specify that, because we did not know how big the room was going to be. The room that was, in the end, allocated to us, did not allow running of experiments with more than 8 participants at the time (see Figure B.1.1 for the room setup).

[10]For an extension of the basic setup where institutional restriction such as *liability* and *verifiability* on the seller's action space are imposed please see Dulleck et al. (2011) or Beck et al. (2014).

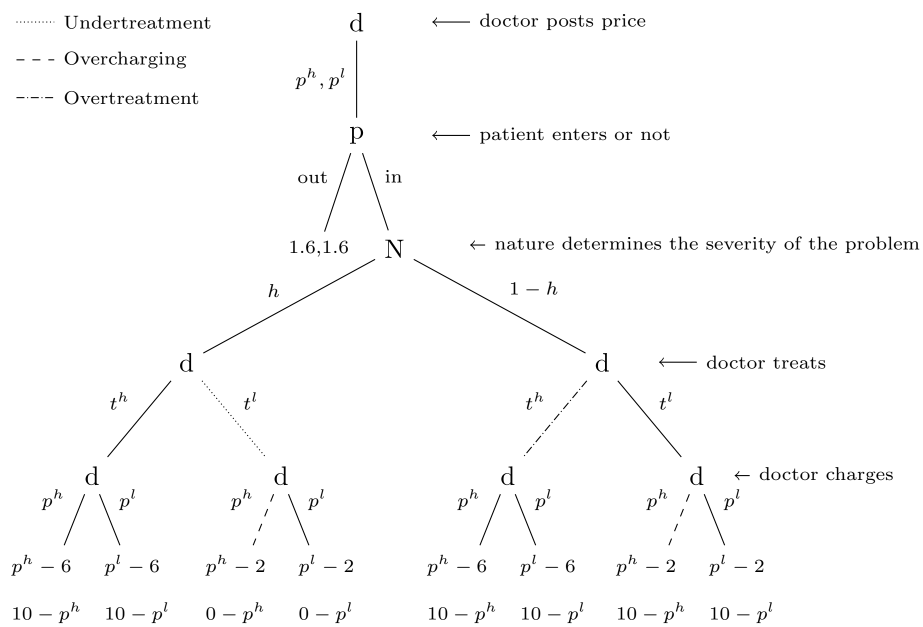


Figure 4.3.1: Game Tree

4.3.2 Theoretical Predictions

Standard economic theory assumes that both sellers and consumers are rational, risk neutral and own-money-maximising agents and that this is common knowledge. A further assumption is that if consumers are indifferent between interacting with a seller and not interacting they decide to participate in the market. Since this is a dynamic game of incomplete information the equilibrium concept applied is a perfect Bayesian equilibrium. The main interest of this study is on symmetric equilibria.[11] Moreover, in markets without institutional restrictions (e.g undertreatment is averted because experts are presumed to be liable) standard economic theory predicts that a doctor would always provide the low treatment and charge the patient for the high treatment under each price vector. By knowing this, the patient would always opt out and not participate in the market if $p^h > 3$. But with such a low p^h even sellers who undertreat

[11]For a theoretical discussion on the solution concept applied please see Dulleck and Kerschbamer (2006).

would earn less than their outside option ($p^h - c^l \leq 3 - 2 = 1 < 1.6 = o$). Hence, no price exists for which both parties would obtain at least the same value as their outside option. Thus, we would expect to observe a market breakdown as described by Akerlof (1970), since undertreatment is a considerable concern in a market with asymmetric information. Additionally, overtreatment and overcharging are going to be an inefficiency concern as well. Hence, the three main inefficiency concerns in a market for credence goods are undertreatment, overtreatment and overcharging.

4.3.3 Experimental Procedure

All experimental sessions were run computerised, using *CORAL* (Schaffner, 2013).[12] The laboratory experiments with a standard participant pool (students) were conducted at the University of Innsbruck in the winter semester 2012/2013.[13] Recruiting for them was organised through ORSEE (Greiner, 2015). A total of 32 participants participated in the experiment.[14] All sessions were started with reading out the instruction to the participants. In the instruction, it was emphasised that all participants receive the same information[15]. The average session length was around 50 minutes. All participants in the experiment received an initial endowment of 10 points to start with. The exchange rate applied in the experiments conducted at the University of Innsbruck was 4 points equal to 1 euro.

The artefactual field experiment was conducted at a large GP's conference (Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin (DEGAM 2014)) in Hamburg. The conference organisers provided us with a room, in

[12]For a detailed introduction of CORAL or to view the homepage that was particularly created for new users, see Schaffner (2013) or `https://github.com/mas802/coral-econ` respectively.

[13]We used the same 32 observation from Chapter 3

[14]We also conducted the same experiments with students from the Queensland University of Technology. However, as we were unable to organise experiments with medical professionals in Australia and due to cultural differences between the participant pools and the medical professionals (e.g. instructions in different languages), we only report in the main text the results of experiments conducted in Germany and Austria. See Appendix Table B.3.2 for the results where we include all the observations.

[15]See Appendix D.1 for the experimental instructions in English. The experiments with students and medical professionals were conducted in German.

which we could set up a mobile lab with eight laptops and one server[16] for the duration of the conference.[17] A total of 40 conference participants took part in the experiment. These sessions were run with eight participants at the time. The payoff structure was also altered to account for the fact that medical professionals have a higher opportunity cost of time. We employed a one-to-one exchange rate, hence, one point equals one euro.[18]

4.4 Results

Table 4.4.1 provides a descriptive overview of the experimental results for medical professionals and students. The last column of the table presents the results of a Wilcoxon Rank-Sum test.

Standard economic theory would have predicted that in a market without institutional restrictions, no trade should take place since sellers would always provide the lower treatment and charge for the higher one. However, this is not what was observed. As depicted in Table 4.4.1, there is a relative frequency of trade of almost 49% in experiments with students and participation in the market for medical professionals is even higher in the experiment at 60%. The difference in trade that was observed between the two participant pools is highly significant at the 1% level. Furthermore, efficiency is also significantly higher ($p < 0.01$) for medical professionals compared to students.

[16]For a picture of the outline and setup of the room please see Figure B.1.1 in the Appendix.

[17]We received great support by setting up a mobile lab from Olaf Bock and his team of the Experimental Laboratory at the University of Hamburg.

[18]We are aware of the fact that an exchange rate of doctors earning 4 times more than students might not be representing an actual real world situation. The values we implemented for doctors are actually on the conservative side, however, considering the financial constraint it was the best we could do. Should there be a bias between the different stake sizes the results with doctors would probably be downward biased.

Table 4.4.1: Overview of Baseline Results Medical Professionals vs Students

	Neutral Framing		
Averages per Period	Medical Prof.	Students	Mean Diff. ($\lvert z - value^{g}\rvert$)
Trade[a]	0.60	0.49	0.11 (3.51)***
Efficiency[b]	0.23	0.05	0.18 (17.78)***
Undertreatment[a,c]	0.50	0.67	-0.17 (3.96)***
Overtreatment[a,d]	0.20	0.11	0.09 (3.01)***
Overcharging[a,e]	0.67	0.83	-0.16 (5.31)***
p^l with trade	3.96	4.33	-0.37 (4.12)***
p^l without trade	4.35	4.90	-0.55 (3.73)***
p^h with trade	7.38	7.06	0.32 (2.84)***
p^h without trade	8.50	8.16	0.34 (2.62)***
Actually charged price	6.62	6.92	-0.30 (1.43)
Profit sellers[f]	2.55	2.82	-0.27 (3.16)***
Profit consumers[f]	1.39	0.52	0.87 (8.91)***
Most prominent price vector	(4,8) 14% (5,8) 8% (6,8) 7%	(6,8) 18% (5,8) 16% (4,8) 7%	
Number of obs. (# participants)	640 (40)	512 (32)	

*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.
[a] relative frequency.
[b] calculated as (actual average profit - outside option) / (maximum possible average profit - outside option)
[c] patient needs t^h, but seller provides t^l.
[d] patient needs t^l, but seller provides t^h.
[e] doctor provides t^l, but charges p^h (with $p^l \leq p^h$ and consumer requiring t^l).
[f] in experimental currency units.
[g] Wilcoxon Rank-Sum Test where Ho: variable$_i(Medical_{prof} == 0)$ = variable$_i(Medical_{prof} == 1)$.

By observing the three main inefficiency variables in Table 4.4.1, undertreatment, overcharging and overtreatment, we detect that medical professionals behave more honestly. They undertreat and overcharge less often than students (both $p < 0.01$). On the other hand, for the main variable of interest in relation to SID, medical professionals overtreat more often than students ($p < 0.01$). However, if medical professionals were only interested in their financial profits, we should also see that they overcharge more often than students, which was not observed. One reason for the higher number of overtreatment occurrences of medical professionals, compared to students, could indicate a problem in a physician's decision heuristics. For instance, medical professionals often cannot know if a diagnosis is correct or if a treatment can cure a patient for certain if it is not a straightforward case. It is therefore often an appropriate solution to prescribe additional treatments to obtain a clearer picture of the patients' medical situation. Additionally, as depicted in Figure 4.4.1 medical professionals, allocated the role of a doctor, earn less than their student counterparts ($p < 0.01$). On the other hand, expert sellers in the experiments conducted with medical professionals receive a higher average payoff than consumers ($p < 0.01$) in experiments conducted with non-professionals. Furthermore, the most prominent price vector chosen by medical professionals is the equal mark-up price[19] (4,8) such that both the doctor and patient would both earn a profit of 2 unless the doctor cheated. In fact, in almost 60% of times doctors behaved honestly and did not choose to undertreat while for students almost 90% of times, if they selected the equal mark-up price, they undertreated the patient. These results combined clearly indicate that medical professionals allocated the role of sellers, in this experiment, do not act in their own self-interests per se; rather, they behave in favour of the consumer. This would be in accordance with a statement made by Blomqvist (1991) and Kessler and McClellan (1996), who suggested that fear of litigation or liability could be one reason why physicians may prescribe more treatments than actually needed, and not to increase their monetary payoff. In line with this result is an observation by Iizuka (2007) who finds that doctors in Japan, who often prescribe and distribute drugs to their patients, react to mark-up differences in the Japanese drug prescription market. In particular, he finds that physicians in Japan tend to care more about out-of-pocket expenses for patients than generating a higher profit for themselves from the mark-up. However, in Figure 4.4.1, it can also be seen that the average payoffs for the consumer in the experiment with medical professionals (1.39) and students (0.52) are below the outside option of 1.6. In fact, half of the participants allocated the role of consumers in the experiment with medical professionals, received more

[19] An equal mark-up price-vector is defined such that $p^h - p^l = c^h - c^l = 4$.

than the outside option of 1.6, while in contrast only two participants playing with students received a higher average payoff than 1.6. Hence, it would have been more beneficial to choose the outside option and stay away from the credence goods market for both participant groups in the respective experiments allocated the roles of consumers. To test the robustness of the results of this study, a multivariate analysis is conducted in the next section.[20]

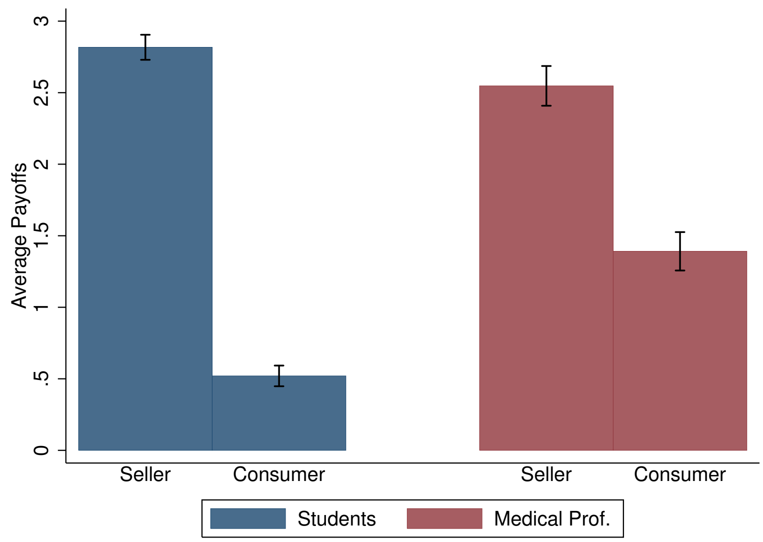


Figure 4.4.1: Average Payoff

4.4.1 Estimating the Effect of Trade, Undertreatment, Overtreatment and Overcharging

Table 4.4.2 presents Probit regression results of medical professionals' impact on the relative frequency of trade, undertreatment, overtreatment and overcharging. The results and the signs of the coefficients of medical professionals are as expected and in line with the analysis in the previous section. Table 4.4.2's specification (1) presents the regression results for the differences in the likelihood

[20]To observe the same analysis for undertreatment, overtreatment and overcharging with individual participant averages, please refer to Appendix B.2. The results have only changed slightly but are still highly significant.

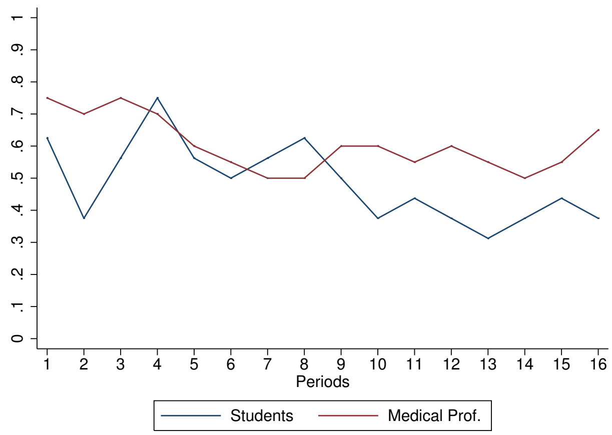
of trade between medical professionals and students. Medical professionals are 14.7 percentage points more likely to participate in the market than students. This is also illustrated by Figure 4.4.2 (a), where, over the 16 periods, the relative frequency of trade of medical professionals is, almost everywhere, higher than for students. This higher decline in trade for students indicates that participants learn about the fact that undertreatment happens and that it is costly for them (Dulleck et al., 2011). This is also illustrated by the significant positive coefficients ($p < 0.01$) of Period 1-3 and Period 4-8 compared to the reference group Period 9-16. In specification (2), the differences between the two participant pools are analysed for undertreatment. A strong and highly significant difference is found at the 1% level. The marginal effect indicates that medical professionals are 26 percentage points less likely to undertreat than students. Furthermore, in Figure 4.4.2 (b) one can depict that for student participants the average rate of undertreatment, over the 16 periods, is in most of the periods above or at equal percentage levels. In three out of 16 periods, however, the frequency of undertreatment was higher for medical professionals. Moreover, it was also observed that the price for the high treatment is statistically significant at the 1% level. The marginal effect is indicating that for every lab$ increase in price, there is a 6 percentage point increase in undertreatment occurrences. The negative coefficients of the period variables are confirming the illustration presented in Figure 4.4.2 (b) where it is seen that the frequency of undertreatment is lower in earlier periods compared to later periods.

Table 4.4.2: Probit Regressions - Baseline

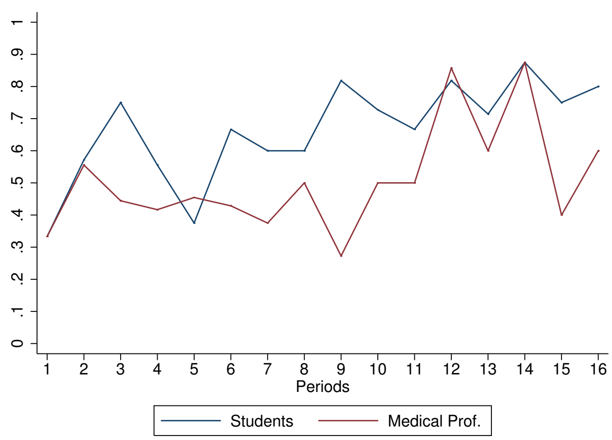
Variables	(1) Trade	(2) Undertreatment	(3) Overtreatment	(4) Overcharging
Medical Prof.	0.375**	-0.693***	0.629***	-1.121***
	(2.27)	(-3.79)	(3.41)	(-6.08)
	0.147	*-0.262*	*0.124*	*-0.248*
Price Low	-0.021	-0.388***	0.196***	-0.570***
	(-0.86)	(-8.25)	(6.06)	(-6.09)
	-0.008	*-0.150*	*0.041*	*-0.128*
Price High	-0.225***	0.162***	-0.004	0.409***
	(-7.17)	(3.57)	(-0.08)	(5.39)
	-0.089	*0.062*	*-0.001*	*0.092*
Period 1-3	0.369***	-0.484**	0.853***	-0.503***
	(3.83)	(-2.55)	(5.32)	(-2.87)
	0.141	*-0.190*	*0.230*	*-0.134*
Period 4-8	0.328***	-0.587***	0.149	-0.195*
	(4.20)	(-4.35)	(1.07)	(-1.65)
	0.127	*-0.228*	*0.032*	*-0.046*
Constant	1.603***	1.385***	-2.521***	0.910***
	(6.27)	(4.06)	(-6.45)	(2.86)
χ^2	70.740***	99.221***	140.982***	85.119***
$Prob > \chi^2$	0.000	0.000	0.000	0.000
(Pseudo) R^2	0.091	0.200	0.142	0.324
N (Clusters)	72	72	72	72
N (Obs.)	1152	550	602	824

Notes: Clustered standard errors, z-statistics in parentheses and marginal effects in italics. The reference groups consist of STUDENTS and PERIOD 9-16. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Specification (3) in Table 4.4.2 compares the likelihood of overtreatment occurrences between medical professionals and students. The coefficient of overtreatment is positive and highly significant at the 1% level. The marginal effect indicates that medical professionals overtreat 12.4 percentage points more often than students. Looking at Figure 4.4.3 (a) no clear difference in trends between the two groups can be observed over time, except for the fact that in most periods, medical experts overtreat more than non-medical experts as indicated in the significant positive coefficient of medical professionals in specification 3. Furthermore, overtreatment over time seems to be higher in earlier periods compared to later ones, as indicated in the significant positive coefficient of Period 1-3. Additionally, for every point increase in the lower price, the tendency to overtreat seems to increase by 4.1 percentage points, while the price for the high treatment appears to have no significant effect at all on the tendency to overtreat.

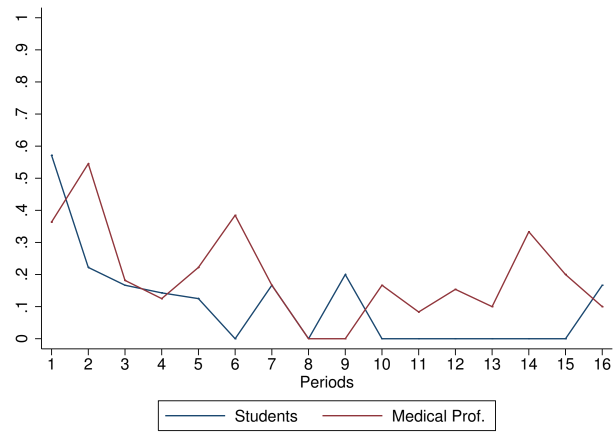


(a) Relative Frequency of Trade

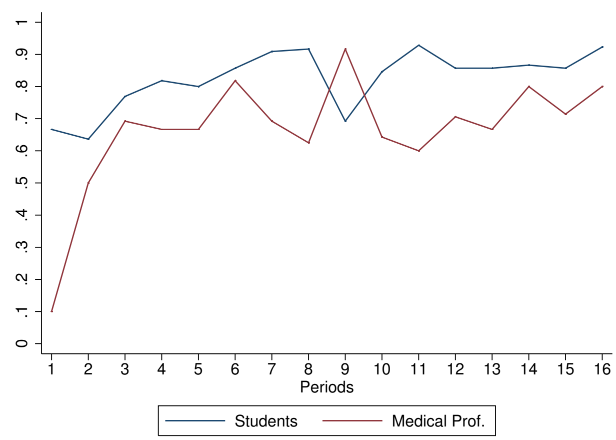


(b) Undertreatment

Figure 4.4.2: Frequency of Trade & Average Rate of Undertreatment



(a) Overtreatment



(b) Overcharging

Figure 4.4.3: Average Rate of Overtreatment & Overcharging

The aim of the last column of Table 4.4.2 is to explore the difference between the two participant pools in their propensity to overcharge. As shown in specification (4) the coefficient for overcharging is significantly different at the 1% level between medical professionals and students. Overcharging is 24.8 percentage points less likely to occur with medical professionals as experts compared to student experts. By looking at Figure 4.4.3 (b) over the 16 periods, we can observe that medical professionals begin with a very low probability of overcharging but their tendency to overcharge increases quickly, confirming the result we obtained for the significant negative coefficient of Period 1-3. Compared to our reference group Period 9-16 overcharging is 13.4 percentage points less likely to occur at the beginning of the experiment. Demographic characteristics such as age and gender were also controlled for, but no significant influence was found and therefore not included in the main text. Please see Appendix Table B.3.1 for the regression results. As further robustness checks we have also conducted a random effects panel regression Table B.3.5, a probit random effects panel regressions Table B.3.3 and the same regression model but with bootstrapped standard errors Table B.3.4. The results do not show any deviations from the original results.

4.5 Concluding Remarks

This paper analysed fraud and service provision with a focus on the SID effect in the health care market by conducting an artefactual field experiment with medical professionals. In particular, the behaviour of university students from IBK, was compared with participants at a large medical GP conference in Hamburg. It was found that medical professionals in an experimental credence goods market overtreat significantly more than students but concurrently undertreat and overcharge significantly less. Hence, the results of this research are similar to those of Beck et al. (2014) who find that car mechanic professionals overtreat significantly more than their student counterparts in the experiment. In line with these results, it is argued that the difference observed is a result of the decision rule learned by car mechanics in the past. We also believe that the difference between medical professionals and students could emerge due to the training and decision heuristics learned by medical professionals through their experience working as doctors. It could also depend on a physicians risk attitude, for instance, when confronted with a straightforward diagnosis; physicians fear of a malpractice lawsuit is negligible, however, when experiencing a complicated incident it is easier for a risk-averse physician to prescribe additional treatments to prevent future complications. Moreover, as previously mentioned, the results in this experiment indicate that medical professionals, who are allocated the role of a doctor, *ceteris paribus*, care more for the patients in the experiments, than do the students. They forgo a higher profit for themselves by choosing three times more often than students the equal mark-up price (4/8). As overtreatment is a frequent occurrence in a market for credence goods when professions such as taxi drivers (see Balafoutas et al., 2013), car mechanics (see Schneider, 2012; Beck et al., 2014) or - as in this paper - medical professionals are analysed, it seems to be of a greater concern than previous experimental papers have suggested (see Dulleck et al., 2011; Kerschbamer et al., 2009).

This research suggests that to tackle the problem of SID in the health care market we need not only to understand the bare numbers of follow up visits and physicians density but also physicians behaviour. Furthermore, our research did not analyse the risk attitude of a physician in particular, but for future research that could be an interesting additional aspect to consider. For example, more risk-averse physicians prescribe more follow-up visits to account for their uncertainty. However, due to the low number of physicians participating in the experiment, it was not possible to run more treatments with institutional restriction such as liability or verifiability. Also, it would have been interesting to observe how physicians behaviour might change when a medical framing was implemented, compared to a neutral framing as in this experiment. In addition, there may also be a selection effect, due to personality traits such as the intellectual curiosity of doctors in this experiment (Slonim et al., 2013). It is not common for doctors to participate in laboratory experiments, as our post survey indicated. More than 50% of doctors have never participated in a laboratory experiment previously. Hence, some participants may have just wanted to find out what the experiment was about and therefore participated, while others, wanted to help us out to obtain a sufficient number of participants. While this research only investigated one area of fraud and service provision, it could also be applied to most situations where asymmetry of information is a key characteristic.

In Chapter 3 it has been analysed whether the concept of being a medical professional affects the behaviour of participants in the experiment (i.e more honestly). The second study in Chapter 4 analysed whether medical professionals behave any different than students in respect to honesty. This raises a question about selection into the medical profession. Are individuals self-selecting into medical professions the same or are they different from other professionals? In the following section, Chapter 5 investigates whether the behaviour of law, accounting and engineering students are any different from the behaviour of medical students, by extending the previous credence goods experiment with some additional games.

Chapter 5

Naturalistic framings and honesty of future professionals: A triple experiment

5.1 Introduction

The average person spends more than 40 years in the workforce, which is a substantial amount of a persons lifespan. Hence, it is essential that individuals allocated themselves in an efficient way to choose the most suitable profession in respect to their talents and aptitudes and to reduce damage to society as a whole. People who are better suited for a job have less stress, feel proud of things they have achieved and feel happier overall than people who are less suitable for the job (Harter and Arora, 2010). Moreover, society benefits from people who are happier or better suited to jobs since their productivity is higher. Hence, self-selecting into a career for which one is not well-matched is not only inefficient for society, but it can also cost society a vast amount of resources. For example, a poor general practitioner (GP) who fails to accurately diagnose a serious illness has the potential to undertreat a patient and in the long run the cost for the treatment is much higher for the patient than it should have been. Furthermore, a GP lacking the appropriate skills, who cannot determine the illness of a patient, possibly prescribes many more treatments than necessary. Therefore, we need people to make better, more informed decisions not only for themselves, but also for society.

Smith (1776) argues that specialisation is one possible solution for society to overcome such problems; the division of labour plays an important part for a society to develop and flourish. For example, in the medical sector, a surgeon who specialises in organ transplant is much better at performing the procedure than a surgeon who only does it sporadically. Specialisation ensures a greater investment in knowledge that is beneficial to society. Becker and Murphy (1994) state that, 'the dependence of specialization on the knowledge available ties the division of labor to economic progress since progress depends on the growth in human capital and technologies' (p. 300). University education is one way to better society where future employees can be educated to obtain the skills needed for the community. For a society to advance, universities play an essential part in industrial innovation through the education of prospective professionals to teach them new and improved techniques or ideas, which they can then apply in their professional lives - for example, public health, infrastructure, justice and the tax system, to name a few. Since individuals are already self-selecting themselves into university courses and thus careers, choosing the appropriate area of study is not only for prospective students a major objective, but also one for society. Companies have to spend a significant amount of money on training to ensure that the person they are going to hire is the right one (Karsan, 2007), and on creating a meaningful work experience for employees as this will decrease the number of people leaving for another job (Scroggins, 2008). If this money is squandered it will evidently harm society's growth in the long run, as companies have less money to invest. Hence, investigating if there are certain desired types of behaviour that are better suited for specific professions is of fundamental importance. Investigating people who are still in the education system (not yet active in the profession) is vital to investigate if the matching, as it currently is, is adequate. Pingle (2010) states that 'history matters because it not only determines the heuristics we have available in our tool box, but it also determines where we are in the evolutionary process of matching heuristics to contexts' (p. 75). Hence, since prospective professionals have not yet learned heuristics that could drive their behaviour in certain situations, investigating students allows us to obtain a clearer understanding of the behaviour of future professionals prior to their practical experience.

Medical professions appear to have a high social status within society and a reputation for honesty and strong ethical beliefs. Society expects from medical professionals an exemplary character. Dishonest and unethical behaviour of doctors can harm society immensely, for example, by overtreating or overcharging patients, which costs the public a significant amount of money. But is this true, do medical professionals behave any better than their peers? From previous research we know that doctors do not always behave as honestly as the they should and, for instance, react to fee incentives (Gruber and Owings, 1996; Gruber et al., 1999; Hughes and Yule, 1992; Jürges, 2007; Fuchs, 1978). This paper will analyse if health, law, engineering and accounting students already possess the traits of professionals as seen by the public in regards to honesty and ethical behaviour, and in particular, whether future medical professionals live up to public expectations of honesty and ethical behaviour when compared to the other prospective professionals. For this paper, it was decided to focus the analysis on the four aforementioned professions because they are all essential to a well-functioning society. Being a lawyer is a very important profession that needs reputable people with an outstanding character. Lawyers are involved in the development of rules and the protection of the interest of the common mass using the condition of the constitution. Ben Letham (2010) nicely states that, 'in a society which is governed by the rule of law, lawyers are a society's gateway to the justice their population is entitled to'.[1] Unethical or dishonest lawyers can take advantage of the law and bend it to their advantage, which might not be in the best interest of society. As an extension, in the legal professions, also corrupt judges, for example, are a risk to society if they let criminal offenders walk free. Another profession that needs people with an immaculate character is accounting. Embezzling money is only one way by which a dishonest accountant can harm society. Minor mistakes in the book keeping journal due to dishonest book keeping can have serious consequences and even lead to the collapse of a company, which is very costly for society. The last professional group in which we are interested are engineers. Engineers are essential for solving and tackling the problems of present and future society. Unethical or dishonest engineers can cause major disasters if they behave in an incorrect way by choosing cheaper and faulty materials, for example, just to increase their profit for a project. The general public view is that people of certain jobs are more honest and ethical than others. While all of these professions require a different skill set, they are, in certain characteristics, quite similar to each other.

Figure 5.1.1 illustrates graphically over time the percentage of people who answered the question how certain jobs rate in regards to honesty and ethics

[1] http://www.goinglegal.com/why-are-lawyers-important-in-our-society-1377884.html, accessed on 3 August 2015.

with either high or very high (y-axis). We can depict that Medical Doctors, Pharmacists and Engineers are rated very similarly, while Accountants receive a less positive judgement, and Lawyers are rated even lower. Furthermore, to obtain a feeling how the aforementioned professions compare to other professions within society, we included a few more occupations in Figure 5.1.2. It can be seen that Medical Doctors, Pharmacists and Engineers are rated higher compared to other jobs. Police Officers are close behind; however, jobs such as Accountants, Lawyers and Bankers are rated lower but considerably higher still when compared to Car Salespeople and Members of Congress. Occupational prestige, as viewed by sociologists, confirms the results from the 2014 Gallup Poll - that medical professionals, in particular Physicians, are looked upon very positively. On the other hand, while Lawyers' public perception is relatively low in regards to honesty and ethics, being a Lawyer is associated with a high occupational prestige within society. Table 5.1.1 presents the occupational prestige scores for the professions of interest.[2]

Table 5.1.1: Occupational Prestige Scores[3]

Professions	1970 Prestige Scores	1989 Prestige Scores
Engineers	67	71
Lawyers	76	75
Pharmacists	61	68
Physicians	82	86
Accountants	57	65

These four professional groups were chosen, not only for their key roles and their benefits to a well-functioning society, but also because they have similarities and differences between the groups that allow us to compare their behaviour with each other. This includes: social status, educational requirements, professional associations and litigation. We are particularly interested in students aka future professionals in these disciplines, since they have not yet collected

[2] `http://ibgwww.colorado.edu/~agross/NNSD/prestige%20scores.html`, accessed on 1 August 2015.

[3] `http://publicdata.norc.org/GSS/DOCUMENTS/BOOK/GSS_Codebook_AppendixF.pdf`, accessed on 1 August 2015.

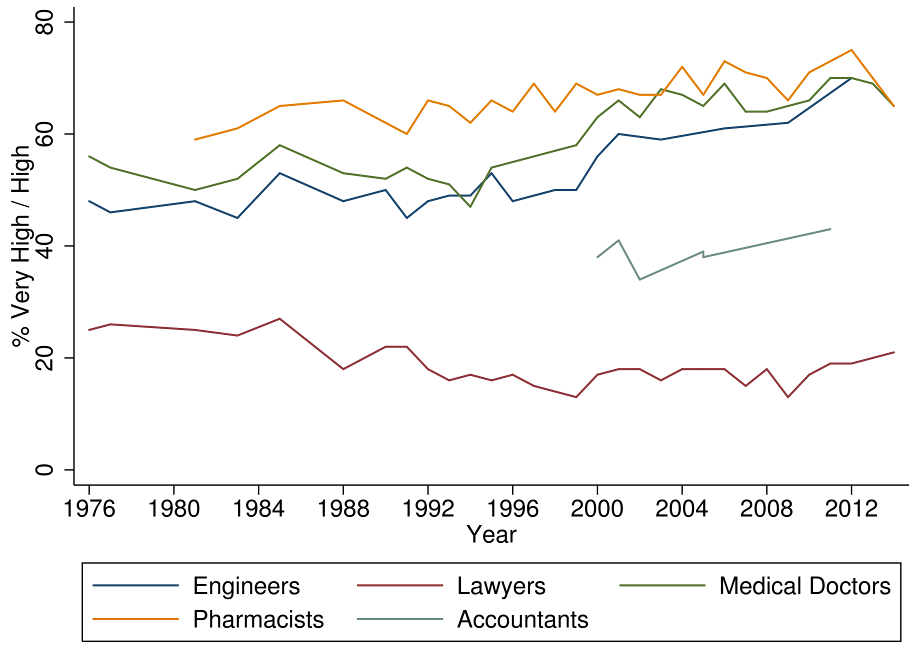


Figure 5.1.1: Honesty/Ethics in Professions
Source: Gallup Poll
(`http://www.gallup.com/poll/1654/honesty-ethics-professions.aspx`,
accessed on 24.07.2015)

any work experience that could bias their behaviour in our experiment. For example, Beck et al. (2014) believe that the differences they observe in their experiment between students and car mechanics are due to learned experiences in the past by professionals confirming Pingle's (2010) previous statement that learned heuristics in the past might matter. Several studies exist that have compared the behaviour of real world experts with students for different environments, and find mixed results; Siegel and Harnett (1964); Dyer et al. (1989); Cooper et al. (1999) find a similar behaviour while Potters and van Winden (2000); Fehr and List (2004); Alevy et al. (2007); Carpenter and Seki (2011) find the opposite.

Conducting a laboratory experiment allows us to observe the behaviour of

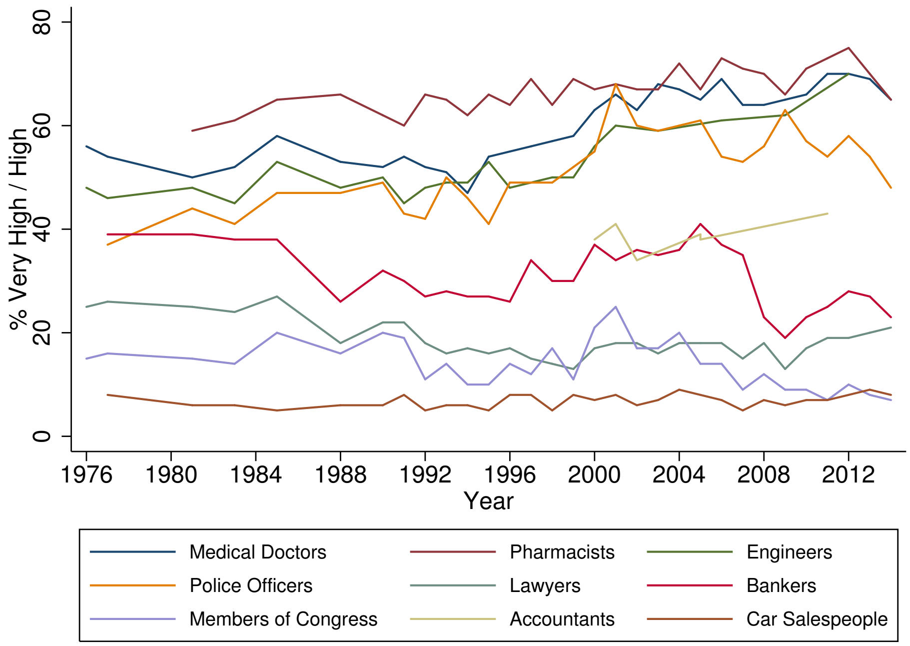


Figure 5.1.2: Honesty/Ethics in Professions - A Comparison
Source: Gallup Poll
(`http://www.gallup.com/poll/1654/honesty-ethics-professions.aspx`, accessed on 24.07.2015)

the different future professional participants in regards to honesty. To the best of our knowledge this is the first study that combines a credence goods experiment with medical framing, a tax compliance experiment with a public goods structure and a self-reporting task with an environmental framing in one single experiment to analyse the behaviour of future professionals. Implementing a different naturalistic framing in each task, that is relevant to at least one particular profession, allows us to obtain a better understanding of the behavioural differences between professions. This is in accordance with Ansink and Bouma (2013, p.3), who state that "it is not clear whether and how subjects in the lab relate to, for instance, a 'business' or 'Wall Street' frame. From this perspective, framing should be matched to a subject pool to which the frame applied".

Specifically, we have a two-player strategic interaction game with type uncertainty and a medical framing relevant for our health subject (see Section 5.2.2), a within-group public goods game with external enforcement risk framed as a tax compliance game relevant for the law and accounting participants (see Section 5.2.4), as well as an honesty experiment with external enforcement risk and an environmental framing applicable for the engineering students (see Section 5.2.3). We find that the participant group for which the framing is relevant behaves more honestly, when compared to a pooled comparison group with the remaining participants, in that particular task with one exception. Law students behave less honestly in the tax compliance framing than the remaining participants. However, most of the results are not statistically significant. Furthermore, comparing every professional group individually against each other, we see that accounting students are overall the most honest group, followed by medical students. The two least honest student groups in our experiment are engineering and law students. The remaining paper is structured as follows, the next section discusses the experimental design, followed by the results and concluding remarks.

5.2 Experimental Design

The experiment is divided into three different tasks. Each task is relevant to a different professional participant group to observe their behaviour in regards to honesty in the experiment. Moreover, every task employs a naturalist framing, matching one of the groups to see if the framing alters the participants' behaviour. In particular, a credence goods experiment with medical framing is run (see Chapter 3 of this document) pertinent to possible future health professionals, a tax compliance game with a public goods structure relevant for law and accounting students and a self-reporting task similar to Friesen and Gangadharan (2013) but with an environmental framing relevant to engineering students.[4]

[4]See Appendix D.1.7 for the experimental instruction.

5.2.1 Experimental Treatments

Given that we have three different tasks in one experiment, we have to control for the order in which we present them to avoid an order effect. Since $3! = 3 \times 2 \times 1 = 6$ we have therefore six possible orders in which we can present the tasks. Moreover, the matching of participants is different for all tasks. While we have no interaction between participants for the self-reporting task with an environmental framing, we have a repeated interaction in groups of four in the tax compliance task. For the credence goods task we do not use a matching group of eight subjects each (4 doctors and 4 patients) in which the subjects are matched after each round within that group as it is normally done; instead, we have that a doctor and a patient are randomly matched with a participant in the experiment of the opposite role after each round. Hence, instead of having a 25% possibility to be matched with the same player of the opposite role, there is only a 12.5% likelihood for this to happen if we run the experiment with 16 participants (8 doctors and 8 patients). However, in all sessions, participants in the roles of doctors cannot identify participants in the opposite role. Furthermore, in the credence goods task we also apply a strategy method for the player in the role of the doctor. Using a strategy method allows us to obtain an observation for the doctor in regards to the price and treatment chosen in every single round, even in the case when the patient decided not to interact with the doctor. The strategy method allows us to collect more data for the person in the role of the doctor, which we otherwise would not have. In either task, participants are required to make a decision in 10 rounds. Hence, participants play a total of 30 rounds.

5.2.2 Two-player Strategic Interaction Game with Type Uncertainty

This task is similar to the credence goods experiment with a medical framing conducted in Chapter 3. We have two roles, a doctor and a patient. The doctor posts prices for the special intravenous drip and the normal intravenous drip. Then the patient decides if they want to be examined by the doctor or not. If the patient chooses yes, then the doctor gets to know with certainty the severity of the disease and administers either the special or normal intravenous drip. If the patient chooses no, then the round ends and both get paid the outside option. However, there are a few differences compared to that outlined in Chapter 3; firstly, we adjusted all the values by a factor of 50 to have a similar payoff structure with the other two tasks of the experiment and secondly, we only run a total of 10 periods instead of 16 periods as in the previous chapters.[5] For a more detailed discussion of the credence goods experimental setup, please refer back to Chapter 3 and/or 4.

5.2.3 Honesty Experiment with External Enforcement Risk

Participants are asked to make a production decision as the owner of a chemical production company. Since this is a risky business there is a chance that a chemical spill can occur. The production decision chosen by the participants directly affects the probability of a chemical spill. Participants can reduce the probability of a chemical spill but it is costly and will therefore reduce their production earnings. Likewise, increasing the probability of a chemical spill will increase a participant's production earnings. Regardless of a chemical spill occurring or not, a participant receives the production earnings. There is, however, a 50% chance that a participant might be inspected. If a participant were to be inspected and a chemical spill has occurred, then a participant has to pay a fine. The probability of having a chemical spill is determined by the computer in accordance with a participant's chosen accident probability and is independent across rounds. Furthermore, a participant is asked to fill in a report about whether an accident has occurred or not. If a participant reports that an accident has occurred, they always have to pay a self-reporting fine, which is less

[5] For example, while in the previous experiment we have a cost of the normal or special intravenous drip of 2 and 6 respectively, for this task we have a cost of 100 and 300 respectively. The same adjustment was applied for the prices the doctor can charge for the different infusions. i.e $I \in \{50, 100, \dots, 550\}$ instead of $I \in \{1, 2, \dots, 11\}$ as in the previous experiment.

than the fine a participant has to pay if they reported they did not have an accident despite having had one and getting inspected.

5.2.4 Within-group Public Goods Game with External Enforcement Risk

In this task, participants are responsible for making a tax declaration decision. Each round represents one year and proceeds in three stages. Additionally, participants are grouped together with four other participants amongst the tax revenue is redistributed. In the first stage of each round, participants receive an income of 100 lab\$, which represents their taxable income. On the amount of income a participant declares, a universal tax rate of 30% will be applied. In the second stage of the experiment, the taxes payable will be deducted from a participant's income. The remaining net income accumulates as wealth. Also in this stage, there is a 10% likelihood that a participant might be selected for an audit, in which the declared income will be checked against the real income of that year. If there is a discrepancy between the real and stated income, a fine of 1.5 times the amount of undeclared income will be deducted from a participant's income of that year on top of all taxes owed. Finally, in the third and last stage of one round, every group member will receive a transfer, which is the sum of tax contributions of all members of the group multiplied 1.6 and divided by the group size (= 4). For instance, if the amount of taxes (i.e. the sum of tax payments of all members of the group) is 30 lab\$, each group member receives a transfer payment of 12 lab\$ ($= 30 \times 1.6/4$).

5.2.5 Experimental Procedure

All experimental sessions were run with CORAL (Schaffner, 2013). Recruiting was done with ORSEE (Greiner, 2015) in Semester 2 2014 and Semester 1 and 2 2015 at the Queensland University of Technology. A total of 108 students from four different majors, namely Health (21), Law (23), Accounting (32) and Engineering (32), participated in the experiment. A total of nine sessions were run. In every session, the tasks were presented in a different order to the participants, to avoid an order effect. The session size varied between 8, 12 and 16 participants in the experiments. At the beginning of the sessions, the instructions were read out to the participants. It was highlighted in the instructions that all participants were to receive the same information. The average session length was approximately 1 - 1.5 hours. In addition to a 5 dollar show-up fee, each experimental tasks associated payoff has been adjusted to obtain an expected payoff of \$5 as well. Hence, the expected total payoff for a participant in the experiment is \$20.

5.3 Results

5.3.1 Credence Goods Task

In the following section, the results for the different experimental tasks are presented and discussed separately. Firstly, the credence goods task with a medical framing is discussed, followed by the self-reporting task with an environmental framing and lastly, the behaviour of the participant in the public goods game with external enforcement risk framed as a tax compliance game.

Since we are mainly interested in how the framing of the different tasks affects the specific student cohort - for instance, for health students, the medical framing - we compare the student cohort for which the frame is relevant against the pooled rest of the students in the experiment. Figures 5.3.1 and 5.3.2 present graphically the means of the different participant pools for our variables of interest. Table 5.3.1 presents the Wilcoxon rank sum test results between the different participant pools and our variables of interest. In the appendix, Tables C.2.1 to C.2.4 present the descriptive statistics of the different variables for the different groups.

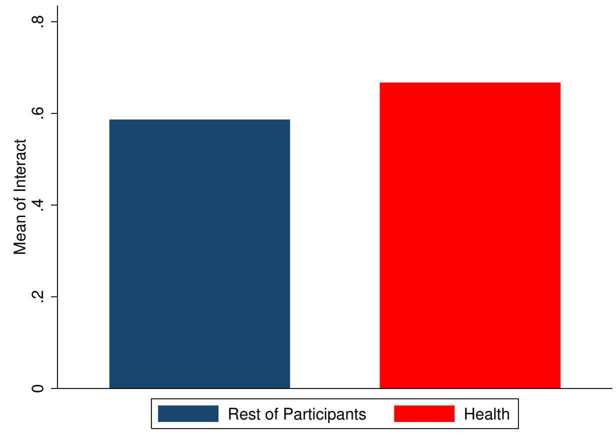
The first four variables, Interact, Undertreatment, Overtreatment and Overcharging, belong to the credence goods task with medical framing. Illustrated in Figure 5.3.1(a), it can be seen that prospective health professionals in the role of the patient, compared to the rest of the participants, choose more often to be examined by and therefore interact with a doctor. As depicted in Table 5.3.1, the difference in the propensity to interact between prospective health professionals and the rest of the participants is significantly different ($p < 0.05$). It seems that prospective health professionals trust the player in the opposite role more than the other participants. This is a desirable result, as medical professions require a lot of soft skills and are intense in human relations. Trusting other people is a sign for a more pro-social behaviour.

Figure 5.3.1(b) depicts the propensity of undertreatment to occur. It can be seen that health students undertreated less, however, no significant difference between the groups can be observed.

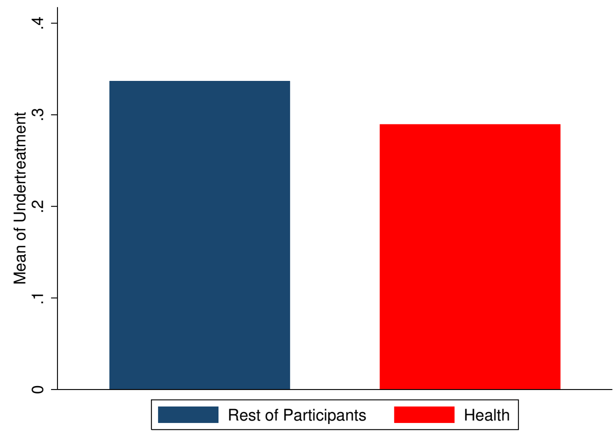
The probability of overtreatment between the two groups is depicted in Figure 5.3.2(a). Health students overtreat slightly more than the rest of the participants but the difference is not significant. The same applies for the propensity to overcharge as depicted in Figure 5.3.2(b). Prospective health professionals overcharge less but the difference is insignificant, as can be seen in Table 5.3.1 when we compare the two means with a Wilcoxon ranksum test.

Figure 5.3.4(a) illustrates the payoffs for health students, acting as doctors in the credence task, and their respective patient.[6] Note that the participants were not aware of the major of the opposing party, so it is meaningless to look at responses of the patients by majors, as they are price takers. We can see that health students in the role of doctors earn, on average, 35 cents less per round than doctors of the other group, however, the difference is not significant (Table 5.3.1). On the other hand, a patient interacting with a prospective health professional earned more than a patient playing against someone from the rest of the participant group. Even though the patient from a health student earned around 25 cents more on average, the difference is not significant, as shown in Table 5.3.1. The striking result is the amount (1.86) that patients from doctors of health majors receive. Furthermore, patients from doctors of the remaining

[6]We have adjusted the values back (dividing by 50) to have similar average profits as in the previous studies.



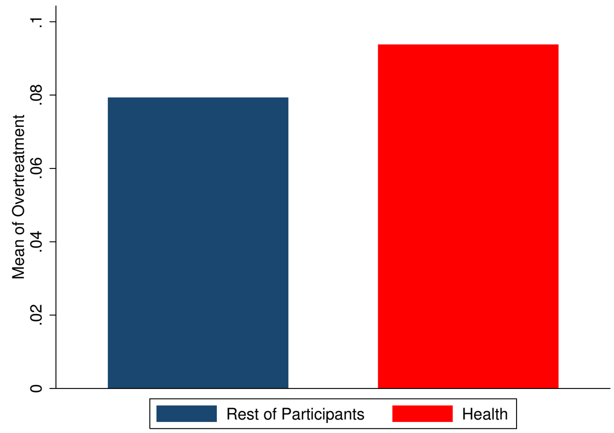
(a) Interact



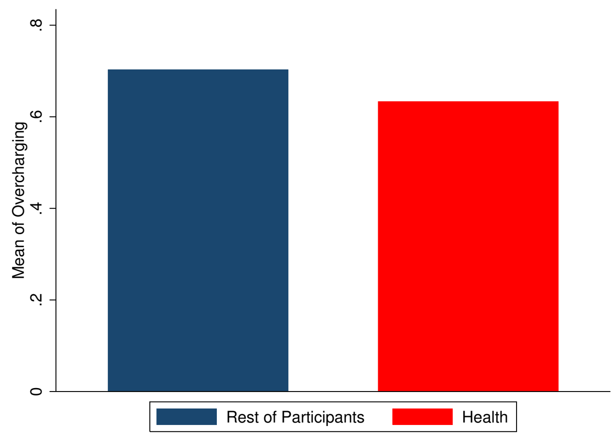
(b) Undertreatment

Figure 5.3.1: Interact & Undertreatment

participants received almost the exact same value (1.61) as people who chose not to interact in the market (1.60). Comparing this with the Dulleck et al. (2011) baseline condition, they did, however, not incorporate a naturalist framing; participants in the comparable role as the patient never received an amount anywhere close to the outside option of 1.6. While economic theory would have predicted a market breakdown as described in Dulleck and Kerschbamer (2006),



(a) Overtreatment



(b) Overcharging

Figure 5.3.2: Overtreatment & Overcharging

section 5.3 on page 29, this is clearly not the case here. Evidently, the medical framing has a pro-social effect on the behaviour of players in the role of doctors when compared to the neutral framing experiment conducted by Dulleck et al. (2011).

Overall, it can be concluded that, in general, the medical framing in the credence goods experimental task affects health students more strongly in the direction of behaving more honestly than the rest of the participant group, however, the effect is not strong enough to obtain a significant difference.[7]

[7]Since we only have 21 prospective health professionals in the experiment this could be one reason why we do not have a strong enough effect.

Table 5.3.1: Overview of Results

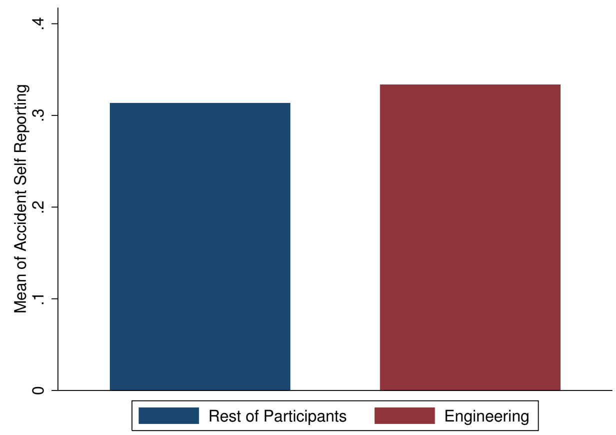
	Credence Goods Task		
Averages per Period	Health	Rest of Participants	Mean Diff. ($\lvert z - value^{f} \rvert$)
Interact[a]	0.67	0.58	**0.08 (2.14)****
Undertreatment[a,b]	0.29	0.34	-0.05 (0.95)
Overtreatment[a,c]	0.09	0.08	0.01 (0.47)
Overcharging[a,d]	0.63	0.70	-0.07 (1.49)
Profit Doctor[e]	2.47	2.82	-0.35 (1.52)
Profit Patient[e]	1.86	1.61	0.25 (1.04)
# of obs. (# participants)	210 (21)	870 (87)	
	Environmental Task		
Averages per Period	Engineers	Rest of Participants	Mean Diff. ($\lvert z - value^{f} \rvert$)
Accident Self-Reporting[a]	0.33	0.31	0.02 (0.53)
Profit	90.27	90.60	0.33 (0.16)
# of obs. (# participants)	320 (32)	760 (76)	
	Tax Compliance Task		
Averages per Period	Law / Acc	Rest of Participants	Mean Diff. ($\lvert z - value^{f} \rvert$)
Tax Compliance Rate[a] (Law)	0.65	0.72	**-0.07 (1.74)***
Profit	109.83	112.27	-2.44 (1.51)
Tax Compliance Rate[a] (Accounting)	0.74	0.69	**0.05 (2.25)****
Profit	112.71	111.34	1.36 (1.33)
# of obs. (# participants) (Law)	(230) 23	(850) 85	
# of obs. (# participants) (Acc)	(320) 32	(760) 76	

*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.
[a] relative frequency.
[b] patient needs t^h, but seller provides t^l.
[c] patient needs t^l, but seller provides t^h.
[d] doctor provides t^l, but charges p^h (with $p^l \leq p^h$ and consumer requiring t^l).
[e] in experimental currency units.
[f] Wilcoxon Rank-Sum Test

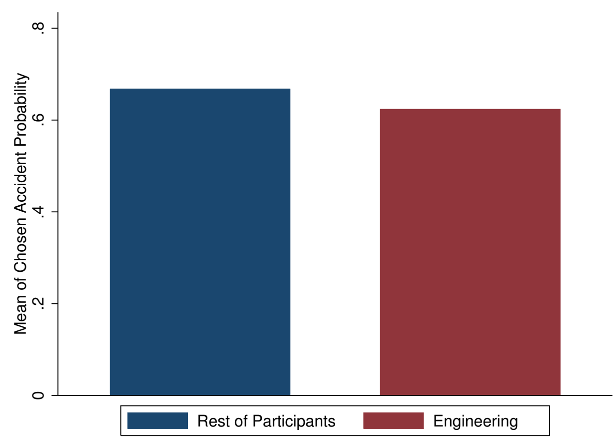
5.3.2 Environmental Task

In this section, analysis is conducted on how the environmental framing in an 'Honesty Experiment with External Enforcement Risk' affects engineering students as the main participant group of interest, compared to the pooled rest of the participants, who self- report an accident and accept a fine more often than the rest of the pooled participant group. However, the average rate of self-reporting an accident is quite low at 33% of times. Nonetheless, this is still much higher than the 15% observed in a similar task conducted in Friesen and Gangadharan (2013). Furthermore, no significant difference between the two participant groups can be detected. In Figure 5.3.3(b), we can also see that engineers seem to be slightly more risk averse due to the fact that, on average, they have chosen a lower accident probability than the pooled comparison group. Both participant groups earned a similar profit (90.27 vs 90.60) on average per round as can be seen in Figure 5.3.4(b) and the insignificant Wilcoxon rank sum test in Table 5.3.1.

Relating these results to the environmental framing of this task, we can see that engineering students behave more honestly i.e more self-reporting than the comparison group. Also the accident self-reporting percentage in our naturalistic framing is more than double the number that Friesen and Gangadharan (2013) obtain in a neutral framing. Moreover, engineering students also choose a lower accident probability and earn a slightly smaller profit than our pooled comparison group. This result indicates that the environmental framing, which is related to the future profession of the participants, has a positive effect on the decision behaviour of future engineers in regards to honesty. However, the results are not significantly different between the two comparison groups.

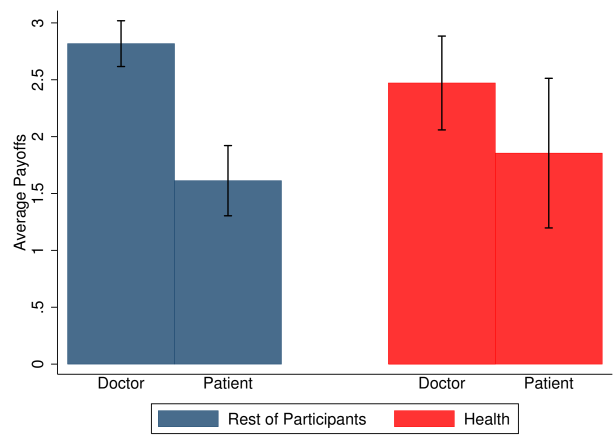


(a) Accident Self-Reporting

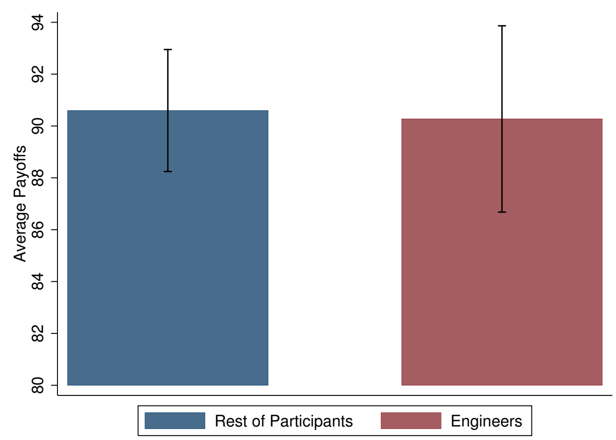


(b) Chosen Accident Probability

Figure 5.3.3: Accident Self-Reporting & Chosen Accident Probability



(a) Payoff Credence Task



(b) Payoff Environmental Task

Figure 5.3.4: Average Payoff Credence & Environmental Task

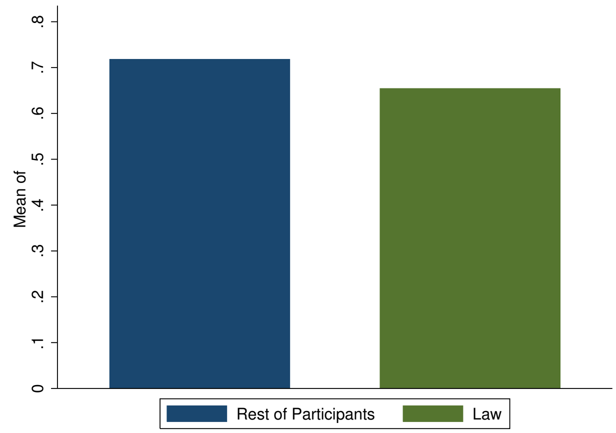
5.3.3 Tax Compliance Task

In this section, the experimental result is analysed for the Within-group Public Goods Game with External Enforcement Risk framed as a tax compliance game. The framing of this task is relevant to two of the invited participant groups, namely for Law and Accounting students. Hence, we have conducted two comparisons. First we compare law students against the pooled rest, which includes accounting students, health and engineering students. In the second comparison, prospective accounting professionals are compared against the pooled rest, which consists of prospective law, health and engineering professionals. Illustrated in Figure 5.3.5 we can depict the tax compliance rate calculated as $\frac{\text{Declared Income}}{\text{Actual Income}}$. A value close to one indicates a high compliance rate while a value close to zero indicates a high propensity to tax evade. On the left side of Figure 5.3.5 we can depict that the tax compliance rate, of law students in comparison with the pooled rest of the participants (0.65 vs 0.72) is lower than the comparison group. Prospective Lawyers' tax compliance rate is in fact weakly significantly lower ($p < 0.10$) than the pooled comparison group as shown in Table 5.3.1. Prospective accountants, on the other hand, behave more compliantly than the pooled rest as can be seen in Figure 5.3.5(b). Prospective Accountants are significantly more compliant ($p < 0.05$) than the comparison group.

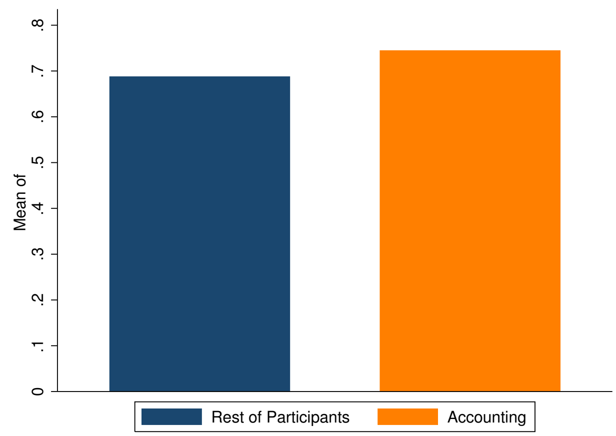
Figure 5.3.6 illustrates the same as the previous graph, however, we have excluded the other future professional group for which the framing is relevant (i.e either law or accounting) from the pooled comparison group. Figure 5.3.6(a) shows the comparison between future law professionals and the pooled rest when we exclude prospective accounting professionals, and Figure 5.3.6(b) illustrates the comparison between prospective accounting professionals and the pooled rest when we exclude future law professionals. The graphs do not change very much.

When analysing the average profit per round of the different groups, we can observe in Figure 5.3.7(a) that law students, despite being less compliant, earn on average less than the comparison group. However, the difference is insignificant as can be seen in Table 5.3.1. Being more compliant was beneficial for accounting students in regards to their average profit per round. On average they earn more than the other group. However, as previously with prospective lawyers, the difference is not significant (Table 5.3.1).

Relating these results to the tax compliance framing and the respective student groups, namely law and accounting, we obtain a very different result for the



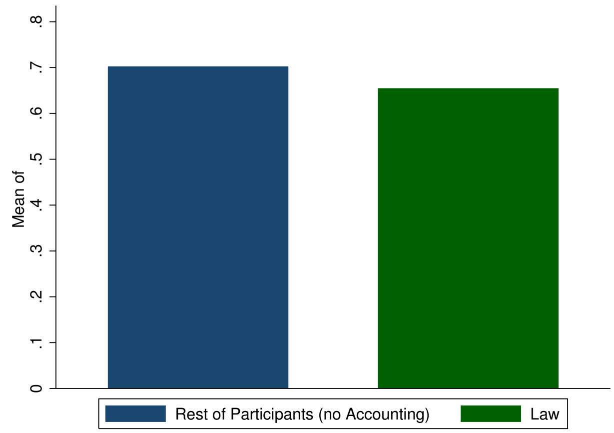
(a) Tax Compliance Rate (Law)



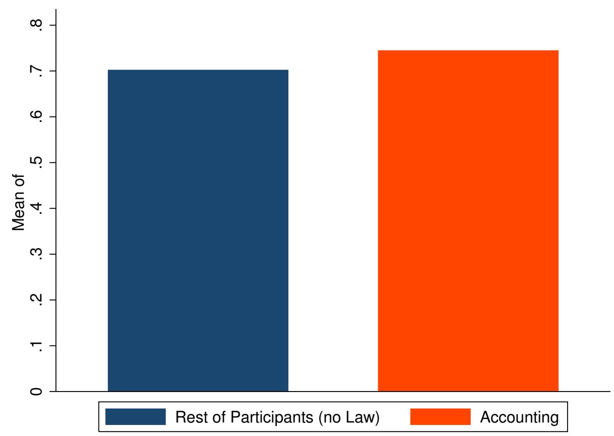
(b) Tax Compliance Rate (Accounting)

Figure 5.3.5: Tax Compliance Rate

two groups. It is found that the tax compliance framing has a stronger influence on compliance for accounting than for law students. The difference between the tax compliance rate of accountants and lawyers is significant ($p < 0.05$), as shown in the Appendix (Table C.1.1). Although it is very important for accountants and lawyers to behave honestly when filling out a tax return, there is a double standard for lawyers. For example, we want honest lawyers that



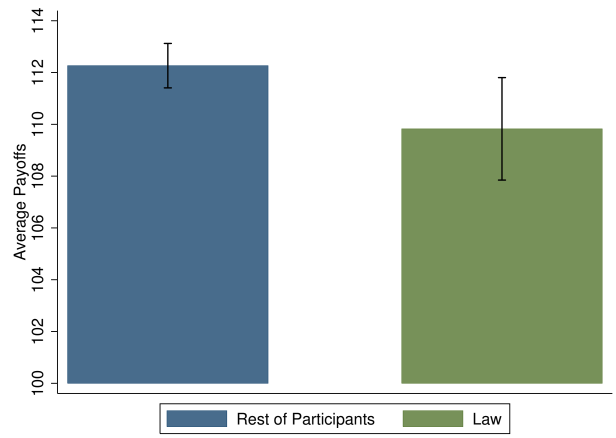
(a) Tax Compliance Rate (Law) no Accounting



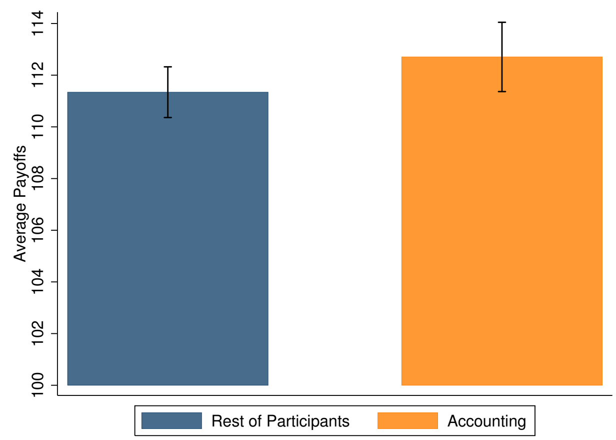
(b) Tax Compliance Rate (Accounting) no Law

Figure 5.3.6: Tax Compliance Rate

are the prosecutors but we also want lawyers that do everything to defend their clients and sometimes this means a lawyer is willing to possibly harm another individual to defend the interest of a client (see Fried, 1976, for a discussion). It might not be as negative that future lawyers did not react as strong to the compliance framing as accountants. On the other hand, the fact that accountants behave significantly more honestly, when the framing is relevant to their future



(a) Payoff Tax Task Law



(b) Payoff Tax Task Accounting

Figure 5.3.7: Average Payoff Tax Task

profession, is indicating that the matching for future accountants is working in the right direction.

To obtain a better understanding of how the different framings in the different tasks affected the different prospective professional groups behaviour, i.e. health workers, engineers, lawyers and accountants, in the experiment, a ranking was developed in which the behaviour of the different groups was rated against each other. In particular, how the different groups behaved in the different task was analysed and the groups were allocated a number between one and four, where 1 = the most honest behaviour and 4 = the least honest behaviour.

Beginning with the inefficiency variables of undertreatment, overtreatment and overcharging from the credence goods task, it can be seen in Figure 5.3.8(a) that engineering students undertreat the most, followed by law and health, while accounting students behave the most honestly. A similar picture emerges when Figure 5.3.8(b) is inspected. The only difference is that law students overtreat the most, followed by engineers and health students. Accountants are again, the group that behaves the most honestly and overtreats the least. Furthermore, the propensity to overcharge illustrates a similar picture, as can be seen in Figure 5.3.9(a). Engineering students overcharge the most, followed by law and then health and again the student group that behaves in the most honest way and overcharges the least are the accounting students. Lastly, this study has a closer look at the payoffs in the credence goods task of patients playing against a student of one of the respective majors allocated the role of the doctor. This study is only interested in payoffs of patients since they are price takers in this task. As depicted in 5.3.9(b), participants in the role of a patient received the highest payoff when they played against health students allocated the role of the doctor. The second highest payoff, by participants allocated the role of patients, was received by playing against law students. The third and the least highest payoff was received by patients who played against doctors from accounting and engineering majors respectively. Summarised in Table 5.3.2 we see that after analysing one task, engineering students are the least honest participants in the credence goods task since they received three times the highest value of four, which is equivalent to having behaved the least honestly, and once the second highest value three.

Moving on to the environmental task and in particular the interest of this research, which is in the accident self-reporting decision. Self-reporting an accident incurred a fine as previously mentioned but for this analysis this is con-

Table 5.3.2: Honesty Comparison Different Major Credence Task

Honesty Variables	Health	Engineering	Law	Accounting
Undertreatment	2	4	3	1
Overtreatment	2	3	4	1
Overcharging	2	4	3	1
Payoff Patient	1	4	2	3
Total Honesty Points	7	15	12	6

1= most honest behaviour, 4 = least honest behaviour

sidered to be the honest behaviour. Figure 5.3.10(a) illustrates the accident self-reporting frequency of the different future professionals. As can be seen, accounting students have the highest frequency of accident self-reporting, followed by engineering, health and the lowest frequency of self-reporting can be observed by law students. Allocating the values one to four accordingly it can be seen in Table 5.3.3 that after two tasks, accounting students are still the most honest student group.

Table 5.3.3: Honesty Comparison Different Major Environmental Task

Honesty Variables	Health	Engineering	Law	Accounting
Undertreatment	2	4	3	1
Overtreatment	2	3	4	1
Overcharging	2	4	3	1
Payoff Patient	1	4	2	3
Accident Self-Reporting	3	2	4	1
Total Honesty Points	10	17	16	7

1= most honest behaviour, 4 = least honest behaviour

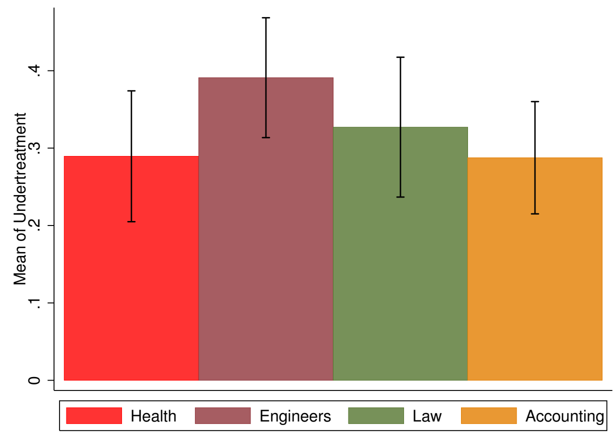
In the last task we are interested in the ratio of declared income over actual income. A value close to one indicates a compliant behaviour. As depicted in Figure 5.3.10(b) we can see that accounting students have the highest tax compliance rate, followed by engineering, health and law students with the lowest rate. Assigning the values for the different positions of the last task we get an overall honesty ranking of the different student groups as illustrated in Table 5.3.4. Overall accounting students received with quite a margin the lowest points for all the different tasks. In other words, accountants behaved the most honestly or more pro-socially compared to the other student groups.

In fact they are the most honest student group in five out of six measures we analysed. The runners up in the honesty ranking are health students, tailed by engineering students and with the least honest student group in this experiment being law students. Clearly the last place for law students is confirming the illustrated Gallup Poll results in Figure 5.1.1 and 5.1.2, how people perceive lawyers in regards to honesty and ethics. In contradiction to the Gallup poll is the result we obtained for accountants and engineers. While prospective accountants behaved much better than the poll would have predicted, engineers students, on the other hand, behaved worse than the public perception would have predicted.

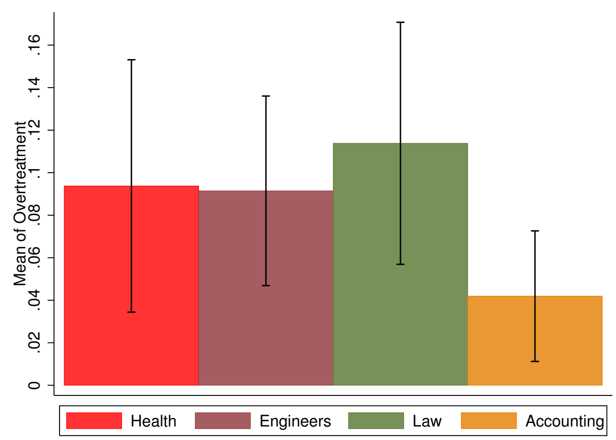
Table 5.3.4: Honesty Comparison Different Majors Overall

Honesty Variables	Health	Engineering	Law	Accounting
Undertreatment	2	4	3	1
Overtreatment	2	3	4	1
Overcharging	2	4	3	1
Payoff Patient	1	4	2	3
Accident Self-Reporting	3	2	4	1
Tax Compliance Rate	3	2	4	1
Total Honesty Points	**13**	**19**	**20**	**8**

1= most honest behaviour, 4 = least honest behaviour

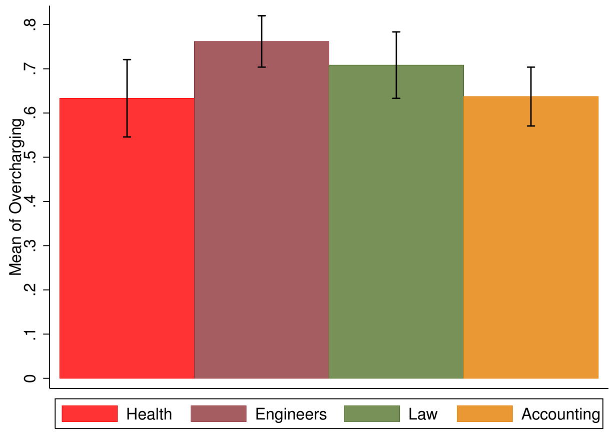


(a) Undertreatment

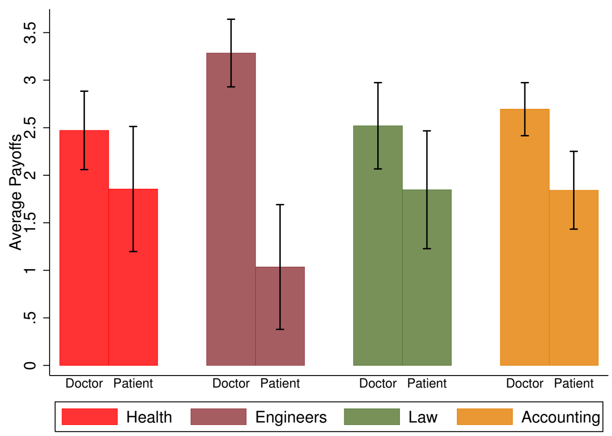


(b) Overtreatment

Figure 5.3.8: Undertreatment & Overtreatment

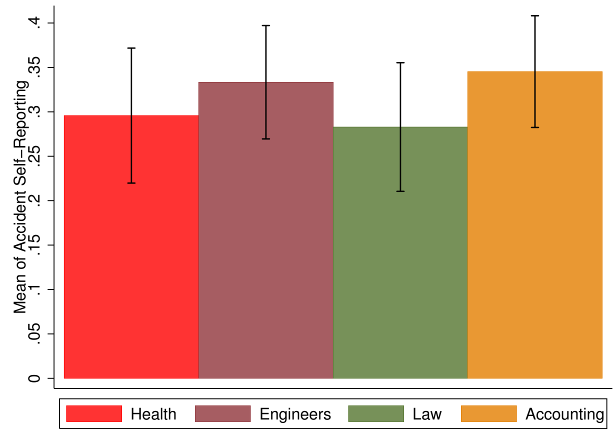


(a) Overcharging

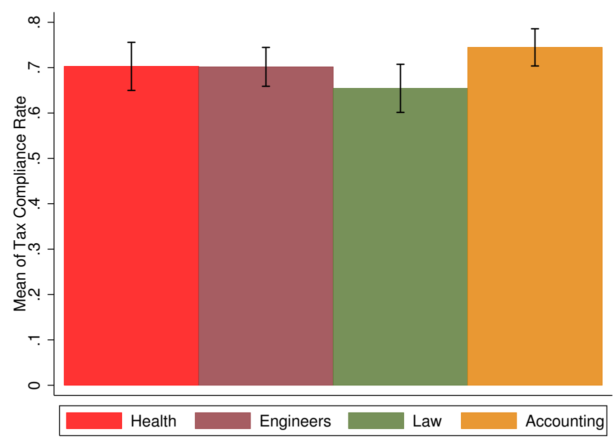


(b) Payoff Credence Goods

Figure 5.3.9: Overcharging & Payoffs



(a) Accident Self-Reporting



(b) Tax Compliance Rate

Figure 5.3.10: Accident Self-Reporting & Tax Compliance Rate

5.4 Concluding Remarks

In this chapter, of interest is the question as to whether individual allocate themselves efficiently, according to their talents and aptitudes, to their most suited profession. To answer this question this researcher has conducted a laboratory experiment with health, engineering, law and accounting students. It is important to observe the behaviour of future professionals prior to them having started working in their chosen profession, due to possible heuristics they would learn, which could influence their behaviour or decisions in our experiment. In particular, this study examined how a naturalistic framing, relevant to one of the aforementioned student groups, affects their behaviour compared to the pooled remaining participants for which the framing is not relevant. Furthermore, to obtain a clearer understanding of the behaviour of the different future professionals in the experiment, a ranking was also developed by comparing their decisions in the different tasks and allocating a value, from one being most honest to four being the least honest accordingly.

Interestingly, it was found that the naturalistic framing affects the behaviour of future professionals, for whom the framing is relevant, more than the pooled comparison group. To be precise, future medical professionals in the credence goods task with medical framing, behave more honestly, i.e. there are less incidences of undertreatment and overcharging. However, in line with the result in Chapter 4, they overtreat more than the comparison group. Furthermore, participants allocated the role of a patient, who played against a future health professional allocated the role of a doctor, earned on average 1.93 points per round. This is much higher than the outside option of 1.6 and a remarkable finding because the patients' payoff in none of the previous credence goods studies that analysed a baseline treatment without a naturalistic framing got anywhere close to the value of the outside option. However, we have to keep in mind that most of the differences are not statistically significant.

In the second task, an environmental framing that is relevant for prospective engineers in our participant pool was implemented. The variable of interest is if participants self-report an accident and incur a fine or do not self-report an accident and get inspected with a certain probability. It was found that engineering students self-report an accident with a higher average frequency than the remaining participants in the experiment. However, the difference is not significant.

For the last task of the experiment, a tax compliance game, with a public goods structure relevant for law and accounting students in our participant pool, was implemented. This study was specifically interested in the ratio of the declared income divided by the actual income. The closer the rate is to one, the more compliant are the participants. Interestingly, for the two groups for whom the framing is relevant, an opposite result was obtained. While law students behave weakly significantly noncompliant than the pooled rest of the participants, accounting students on the other hand, behave significantly ($p < 0.05$) more compliant than the comparison group.

In addition, a simple ranking was also developed to obtain a clearer understanding of how the different future professional groups perform, compared to each other. For this experiment, it was found that accounting students behave much more honestly than the rest of the professionals in all but one observation. Health students follow in the second place, while engineering and law students occupy the two last positions in this ranking. While for law this is not surprising - at least when we compare it to the Gallup poll question for honesty and ethics in professions (see Graph 5.1.1) - for engineering this is unexpected. Overall, we can conclude that the self-selection for professions such as accounting and health works, but, for engineering and law professions it does not entirely and needs more investigation.

However, since these participant numbers are rather low for future health (21) and law (23) professionals, the results have to be taken with caution; additionally, with such low participant numbers no policy implication can be outlined as the results are not robust enough to warrant any suggestion to policy makers.

Chapter 6

Concluding Remarks

6.1 Summary of Findings

This body of work has added to the experimental literature, specifically in credence goods. The innovation in this work, is derived from the implementation of a naturalistic or medical framing, and the utilisation of medical professionals as participants in the experiment. In this thesis, inefficiencies were analysed in the healthcare market by conducting a series of economics experiments, primarily using the credence goods game. It investigated how a naturalistic or neutral framing influenced the honest behaviour of different participant pools. This work is divided into three studies, which addressed different aspects of the forces driving healthcare costs. Together, all three studies contributed to the existing literature on credence goods, framing, artefactual field and laboratory experiments. The healthcare market is characterised by an information asymmetry of an expert seller (doctor) and a consumer (patient). The medical sector is probably one of the most important sectors that relies highly on expert services. Credence goods is the term in economics that describes a situation in which the expert seller knows more about the quality of a good a customer needs than the customer herself (Dulleck and Kerschbamer, 2009). Such asymmetries of information between sellers and consumers may give rise to inefficiencies, such as undertreatment, overtreatment and overcharging, and are a major concern. Unethical experts, for example, can take advantage of this knowledge superiority and abuse their power by providing more treatments than are necessary or charging a higher price for a treatment they have performed. This research is particular interested in the behaviour of the participants in the experiments in regards to honesty. For a society to be functioning well, it is essential that dishonest and unethical behaviour of experts is virtually non-existent.

Each of the studies included in this thesis utilised the economics of credence goods and specifically credence goods laboratory experiments to analyse these inefficiencies, as previously mentioned. In Chapter 3, which served as a baseline for the following Chapters, the effect of a naturalistic (medical) framing and its impact on the behaviour of a common student participant pool in regards to honesty was determined. Chapter 4 extended the analysis in Chapter 3 by examining the behaviour of medical professionals in a credence goods experimental setting with a neutral framing. Finally, Chapter 5 complements the two previous studies and translates them into a larger framework. Different future professional groups were included and their behaviour analysed in regards to honesty in different naturalistic framed tasks.

This final chapter is organised as follows. The reminder of this section briefly summarises the studies and outlines the innovation for each. In the following section, the shortcomings of each study are discussed, followed by policy implication and finally, the thesis concludes with a section for further research directions.

In Chapter 3, which serves as a baseline for the following chapters, the focus is on whether participants behave differently when we apply a naturalistic framing. We know from psychology that it is important how something is presented. If the framing of an experiment is relevant to a situation that participants can relate to, does it change their behaviour. The innovation of this chapter is the introduction of a medical framing in a credence goods experimental setting. Specifically, interest was in how the framing impacts on the behaviour of students in relation to the trade, undertreatment, overtreatment and overcharging. While the experiment conducted with a neutral framing is a replication of Dulleck et al. (2011), the results indicate that the participants in the baseline treatment with a medical framing behave much more honestly than in the neutral framing. For instance, undertreatment was 45 percentage points less likely to occur in medical framing compared to the neutral framing. However, when an institutional restriction such as liability or verifiability was implemented, as strong an effect of medical framing compared to neutral framing, as in our baseline treatment, was not found. Since with the introduction of liability the action space of players is limited and undertreatment is impossible, no differences or even differences that go the wrong direction, between medical and neutral framings, can be found. However, medical framing also has a minor or no effect when we introduce verifiability, which is a bit of a surprise. These mixed results for verifiability suggest that more research into the implementation of naturalistic framings are needed in a credence goods experimental setting.

In Chapter 4, the analysis of the previous chapter was extended by examining the behaviour of medical professionals in a credence goods experimental setting. SID or overtreatment is often seen as a major contributor to the increasing cost in healthcare expenditures. As far as it is known, this is the first study that has analysed the behaviour of medical professionals by conducting an artefactual field experiment in a credence goods setting at a large medical conference in Germany. Of particular interest was medical professionals propensity to overtreat, or in other words, their incentives to induce demand. Specifically, the baseline treatment in a neutral framing was conducted and compared to the results that were obtained previously in Chapter 3 with students. It was observed that medical professionals undertreat and overcharge significantly less, however, their propensity to overtreat is significantly higher than for the student reference group. The findings in this chapter are in line with Beck et al. (2014), who find that car mechanic professionals overtreat significantly more than their student counterparts in their experiment. One possible explanation for this finding is a learned decision heuristics of medical professionals in situations of uncertainty to describe more treatments to prevent future legal proceedings against them if medical complication should arise. Beck et al. (2014) also argue that the difference they observe is a result of the decision rule learned by car mechanics in the past.

In the first two studies (Chapter 3 and 4) we have analysed the behaviour of a standard participant pool in a credence goods experiment with medical and a neutral framing and compared it to the behaviour of medical professionals from a large conference in Germany in a neutral framed, credence goods experiment. Chapter 5 complements these two studies and translates them into a larger framework. In particular, it was of interest whether society allocates people in an efficient way, such that people choose their most suitable profession according to their aptitudes and talents required for that job. Furthermore, since these are professions that occupy important positions within society, an honest and ethical behaviour should be an essential quality of the people working in it. Do future professionals behave more honestly when the framing of the experimental task is relevant to their future professions? To be able to answer this question, a laboratory experiment was conducted with future health, engineering, law and accounting professionals. To analyse prospective professionals prior to their work experience is an advantage, as they have not yet learned heuristics that could bias their behaviour. The innovation of this study is the utilisation of three different tasks in one experiment, of which each task has a naturalistic framing relevant to one of the prospective professional groups. All tasks observe honesty in a different environment and framing. What was found is that, in the credence goods task with medical framing, future health professionals behave more honestly, i.e. less undertreatment, overtreatment and overcharging, than the pooled comparison group consisting of the other professions. Furthermore, the payoff paid to the patient by a prospective health professional allocated the role of a doctor in the credence goods experiment is much higher (1.93) than the outside option of 1.6. Comparing the results of the patient's average payoff (1.93) in medical framing from Chapter 5 to the patient's average payoff (1.20) in Chapter 3, it can be seen that future health professionals allocate a much higher payoff to the patient than participants from a normal student pool. Analysing the task and framing that is particularly designed for future engineers in our participants pool, it is found that engineers self-report an accident, despite having to pay a fine for sure, more often than the pooled comparison group. For the last two professional groups, future lawyers and accountants, a tax compliance experiment with a public goods structure is included. The results are mixed. While future lawyers behave significantly less compliantly than the comparison group, prospective accountants on the other hand behave significantly more compliantly. Although it is very important for accountants and lawyers to behave honestly when filling out a tax return, there is a double standard for lawyers. For example, we want honest lawyers that are the persecutors but we also want the lawyers that do everything to defend their clients and sometimes this means a lawyer is willing to possibly harm another individual to defend the interest of a client (see Fried, 1976, for a discussion). Therefore, it might not be as unexpected that future lawyers did not react as strong to the compliance framing as accountants. On the other hand accountants behave significantly more honestly, when the framing is relevant to their future profession, which is indicating that the matching for future accountants is working in the right direction.

Moreover, when a ranking was developed in relation to the decisions of the different future professional groups in the different tasks, it could be seen that future accountants behave the most honestly in comparison to all other groups, followed by future health and engineering professionals. The least honest prospective professional group are the lawyers in our experiment, which is in line with the public perception expressed in the Gallup Poll, where lawyers received the lowest rating for the groups included in this study[1] However, since this research only has a very limited number of observations, the results have to be taken with caution.

[1] `http://www.gallup.com/poll/1654/honesty-ethics-professions.aspx`, accessed on 24.07.2015.

6.2 Policy Implications

Part of the increasing cost of healthcare is driven by doctors behaviour learned in training. For major treatments or difficult diagnosis, having a policy of forced second opinion could lower overtreatment. The cost of obtaining a second opinion is much more likely to be significantly lower than the cost of unnecessary treatments. By doing this it would, for doctors, reduce the fear of litigation, since there is a second opinion, which would also reduce the need to overtreat. As a result patients are better off (second opinion catches more misdiagnosis), doctors are better off (less litigation and better diagnosis), and lastly it lowers the burden on the health care system by reducing costs.

6.3 Shortcomings

In this section, the main shortcomings of each study which have been discussed in the respective chapters, are outlined. An important shortcoming in Chapter 3 is the fact the no medical students were recruited as participants in the experiment. It would have been interesting to observe, if medical students behave differently compared to a standard participant pool. With respect to Chapter 4, the main limitation is the rather low participant number in the experiment.[2] It makes it difficult to formulate policy implications because low participant numbers normally imply results that lack robustness. Increasing the participant numbers will not only help to obtain a more robust result, it will also help to increase the probability that what we observe is indicative of the larger population. Furthermore, in Chapter 4 the study was unable to conduct a naturalistically framed experiment with medical professionals that would have given it a better understanding and a nice comparison of the behaviour between

[2] Recruiting participants at a medical conference in Germany turned out to be much harder than expected. The conference sessions were very well attended and not many participants were wandering around that we could have approached. Also, the design of the experiment did not help as we needed at least 8 people to concurrently participate in the experiment. While this is not a problem when running experiments in a laboratory at a university it is a large flaw when running experiments at a conference.

medical professionals and a student subject pool in such framed experiments. An important limitation in Chapter 4 is related to the fact that only the behaviour of medical professionals from one particular country with one medical pay scheme was analysed. It might well be the case that different schemes in different countries foster a different behaviour of physicians and therefore these conclusions or recommendations would not apply. Furthermore, running a reputation treatment similar to Dulleck et al. (2011)would have been closer to a real world interaction between a doctor and a patient as well. A further caveat might be the fact that a risk aversion task has not been included in the questionnaire. It could have helped to explain a bit more the higher overtreatment results of medical professionals as the SID problem and the fear of litigation of physicians for malpractice might be related to their risk attributes. Moreover, the artificial environment of the lab combined with an abstract framing could also be a problem for the behaviour of medical professionals.

In Chapter 5, one of the main limitations is a similar problem to that previously discussed in Chapter 4. The low participant numbers in the experiment does not allow for policy implication to be formulated or drawn from.[3] Furthermore, QUT also does not have a medical college, hence, the subject pool does not consist of prospective doctors. However, since doctors are the main focus of this thesis and the people of most interest, it is difficult to derive conclusions and extrapolate them to a wider medical audience.

6.4 Further Research

The credence goods inefficiency problem of undertreatment, overtreatment and overcharging, affects every country around the globe. However, most, if not all, credence goods experiments were conducted in a German speaking country. Hence, a logical extension of Chapter 3 would be to conduct experiments in different countries as it would be interesting to conduct country comparisons

[3] Not using the normal business student participant pool created some problems. The recruitment of medical and law students turned out to be the biggest challenge for this part of my thesis.

with different student participant pools to see if participants in other countries behave similarly or not. Oosterbeek et al. (2004) conducted a meta-analysis on ultimatum game experiments. In 37 papers from 25 different countries they analysed, they find no statistical differences between the behaviour of the proposers, however, the behaviour of the responders across regions is different. Hence, investigating whether the credence goods dilemma is similar across or more prevalent in certain countries than in others, is important when formulating policies that are addressing a countrys particular circumstance. Furthermore, the same experiments could be run with medical students to investigate if a naturalistic framing relating to their chosen professions has a larger influence (i.e. more honest) on their behaviour when compared to a standard participant pool.

The analysis in Chapter 4 of this thesis allows for a number of extensions. The most straight forward, but not necessarily the easiest, is to increase the participant number in the experiment, as we only have 40 participants from Germany. Since the experiment run was in a neutral framing and the doctors could not really relate the context of the experiment to familiar situation it cannot be positively ascertained that the behaviour of medical professionals would be the same when a medical framing was introduced. For example, Cooper et al. (1999) found that the context framing influenced the behaviour of managers much more compared to students than the neutral framing. Moreover, Cohn et al. (2014) found that banking professionals behave less honestly if the framing of the experiment relates to a financial environment compared to a neutral framing. They believe that this is due to the culture that has currently taken place in the financial sector. Hence, it might well be the same for doctors that if the framing relates to a familiar situation, they behave differently than in the neutral framing conducted here. In addition, as the SID problem in the healthcare sector is not only relevant to doctors, it could be interesting to broadening the base of healthcare professionals. To investigate dentists would be a logical and interesting extension due to the similarities with physicians and the characteristics of the expert patient relationship. The dentist-patient relationship might even suffer more from inefficiencies such as overtreatment and overcharging than the doctor patient interaction. Moreover, it would also be interesting to investigate doctors' risk attitudes. A risk-averse doctor might prescribe more treatments when the diagnosis is not straightforward due to the fear of litigation, while a risk-seeking doctor might not care much. Running a credence experiment with a Holt and Laury (2002) task in the post survey of the experiment or a task designed by (Dohmen et al., 2011), which is specifically relevant in a medical domain, would be a logical extension of this study as the risk attitudes of the doctors could be easily compared to the overtreatment decision made in the experiment. Finally, as many countries have different medical pay schemes, running the same experiment in different countries could be interesting, as different medical systems may foster different behaviour of doctors.

Similarly, Chapter 5 of this thesis also offers numerous extensions which could be interesting to pursue. The first and main point would be to increase the number of subjects in the experiment, as we only had 21 health, 23 law, 32 accountants and 32 engineering students that participated. Answering the question if society allocates the people correctly is an important question that needs to be addressed. It can cost society a vast amount of resources and money if this is not efficiently executed. Hence, extending the participant number is only the beginning. More professions that occupy important positions within society should be included in future studies. For example, dishonest bankers can harm society significantly, as the recent global financial crisis has shown. Furthermore, the wording of the tax compliance task could be changed in a future experiment to better emphasise the interaction between an expert and client. Instead of writing *taxes payable on your declared income* this could have been phrased such that the participants in the experiments are tax professionals who have to fill out a tax return for a client to incorporate the credence goods dilemma further and go towards the direction of the study conducted by Balafoutas et al. (2015a), which investigated tax compliance utilising a credence goods experiment. A further extension could also be to narrow this down to only health students. Would we observe a different behaviour of prospective doctors compared to nurses, paramedics or pharmacists - as incentives for pharmacists might be completely different than incentives for nurses - for example?

Bibliography

Ahlert, M., Felder, S., and Vogt, B. (2008). *How Economists and Physicians Trade Off Efficiency and Equity in Medically and Neutrally Framed Allocation Problems?* Jacobs Univ., FOR 655.

Akerlof, G. A. (1970). The market for "Lemons": quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Al-Ubaydli, O. and List, J. A. (2013). On the generalizability of experimental results in economics: with a response to camerer. *NBER Working Paper 19666*.

Alevy, J. E., Haigh, M. S., and List, J. A. (2007). Information cascades: Evidence from a field experiment with financial market professionals. *The Journal of Finance*, 62(1):151–180.

Alger, I. and Renault, R. (2007). Screening ethics when honest agents keep their word. *Economic Theory*, 30(2):291–311.

Andreoni, J. (1995). Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. *The Quarterly Journal of Economics*, 110(1):1–21.

Ansink, E. and Bouma, J. (2013). Framed field experiments with heterogeneous frame connotation. *MPRA Working Paper 43975.*

Ariely, D. (2008). Predictably irrational: the hidden forces that shape our decisions. *New York. NY, Etats-Unis: HarperCollins Publishers.*

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American economic review*, 53(5):941–973.

Balafoutas, L., Beck, A., Kerschbamer, R., and Sutter, M. (2013). What drives taxi drivers? a field experiment on fraud in a market for credence goods. *The Review of Economic Studies*, 80(3):876–891.

Balafoutas, L., Beck, A., Kerschbamer, R., and Sutter, M. (2015a). The hidden costs of tax evasion.: Collaborative tax evasion in markets for expert services. *Journal of Public Economics*, 129:14–25.

Balafoutas, L., Kerschbamer, R., and Sutter, M. (2015b). Second-degree moral hazard in a real-world credence goods market. *Forthcoming in the Economic Journal*.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *The American Economic Review*, 97(2):170–176.

Beck, A., Kerschbamer, R., Qiu, J., and Sutter, M. (2014). Car mechanics in the lab – investigating the behavior of real experts on experimental markets for credence goods. *Journal of Economic Behavior & Organization*, 108:166 – 173.

Becker, G. S. and Murphy, K. M. (1994). The division of labor, coordination costs, and knowledge. In *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*, pages 299–322. The University of Chicago Press.

Bester, H. and Dahm, M. (2014). Credence goods, costly diagnosis, and subjective evaluation. *CEPR Discussion Paper No. DP10254*.

Bickerdyke, I., Dolamore, R., Monday, I., and Preston, R. (2002). Supplier-induced demand for medical services. *Canberra: Productivity Commission Staff Working Paper*.

Blomqvist, Å. (1991). The doctor as double agent: Information asymmetry, health insurance, and medical care. *Journal of Health Economics*, 10(4):411–432.

Bock, O., Nicklisch, A., and Baetge, I. (2012). hroot: Hamburg registration and organization online tool. *H-Lab Working Paper (1)*.

Bolton, G. E. and Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *The American Economic Review*, 90(1):166–193.

Brandts, J. and Charness, G. (2003). Truth or consequences: An experiment. *Management Science*, 49(1):116–130.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., and Wiesen, D. (2013). How effective are pay-for-performance incentives for physicians? a laboratory experiment. Ruhr economic papers 2013-413.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., and Wiesen, D. (2014). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. HERO On line Working Paper Series 2014:3, Oslo University, Health Economics Research Programme.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Carpenter, J. and Seki, E. (2011). Do social preferences increase productivity? field experimental evidence from fishermen in toyama bay. *Economic Inquiry*, 49(2):612–630.

Charness, G. (2010). Laboratory experiments: Challenges and promise: A review of theory and experiment: What are the questions? by vernon smith. *Journal of economic behavior & organization*, 73(1):21–23.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Cohn, A., Fehr, E., and Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature*, 516(7529):86–89.

Cooper, D. J., Kagel, J. H., Lo, W., and Gu, Q. L. (1999). Gaming against managers in incentive systems: Experimental results with chinese students and chinese managers. *American Economic Review*, 89(4):781–804.

Darby, M. R. and Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of Law and Economics*, 16(1):67–88.

Deutsch, M. (1958). Trust and suspicion. *The Journal of Conflict Resolution (pre-1986)*, 2(4):265–265.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2):459–478.

Dulleck, U., Gong, J., and Li, J. (2015). Contracting for infrastructure projects as credence goods. *Journal of Public Economic Theory*, 17(3):328–345.

Dulleck, U. and Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*, 44(1):5–42.

Dulleck, U. and Kerschbamer, R. (2009). Experts vs. discounters: Consumer free-riding and experts withholding advice in markets for credence goods. *International Journal of Industrial Organization*, 27(1):15–23.

Dulleck, U., Kerschbamer, R., and Konovalov, A. (2014). Too much or too little? price-discrimination in a market for credence goods. Working Paper 2014-13, Faculty of Economics and Statistics, University of Innsbruck.

Dulleck, U., Kerschbamer, R., and Sutter, M. (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *The American Economic Review*, 101(2):526–555.

Dyer, D., Kagel, J. H., and Levin, D. (1989). A comparison of naive and experienced bidders in common value offer auctions: A laboratory analysis. *The Economic Journal*, 99(394):108–115.

Eckel, C. and Gintis, H. (2010). Blaming the messenger: Notes on the current state of experimental economics. *Journal of Economic Behavior & Organization*, 73(1):109–119.

Ellingsen, T., Johannesson, M., Mollerstrom, J., and Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, 76(1):117–130.

Ellingsen, T., Johannesson, M., Tjøtta, S., and Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, 68(1):95–107.

Emons, W. (1997). Credence goods and fraudulent experts. *The RAND Journal of Economics*, 28(1):107–119.

Emons, W. (2000). Product differentiation and price competition between a safe and a risky seller. *Journal of Institutional and Theoretical Economics*, 156(3):431–444.

Fehr, E. and List, J. A. (2004). The hidden costs and returns of incentivestrust and trustworthiness among CEOs. *Journal of the European Economic Association*, 2(5):743–771.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Folland, S., Goodman, A. C., and Stano, M. (2013). *The economics of health and health care*. Pearson, Upper Saddle River, N.J, seventh edition.

Fong, Y.-f. (2005). When do experts cheat and whom do they target? *The RAND Journal of Economics*, 36(1):113–130.

Frey, B. S. (1997). *Not Just for the Money. An Economic Theory of Personal Motivation*. Edward Elgar, Cheltenham.

Fried, C. (1976). The lawyer as friend: The moral foundations of the lawyer-client relation. *Yale Law Journal*, 85(8):1060–1089.

Friesen, L. and Gangadharan, L. (2013). Designing self-reporting regimes to encourage truth telling: An experimental study. *Journal of Economic Behavior & Organization*, 94:90–102.

Fuchs, V. R. (1978). The supply of surgeons and the demand for operations. *The Journal of Human Resources*, 13:35–56.

Gächter, S., Orzen, H., Renner, E., and Starmer, C. (2009). Are experimental economists prone to framing effects? a natural field experiment. *Journal of Economic Behavior & Organization*, 70(3):443–446.

Glazer, J. and McGuire, T. G. (1996). Price contracts and referrals in markets for services. Technical report, Tel Aviv University, Faculty of Management, The Leon Recanati Graduate School of Business Administration.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.

Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization*, 106:367–378.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Grether, D. M. and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, 69(4):623–638.

Gruber, J., Kim, J., and Mayzlin, D. (1999). Physician fees and procedure intensity: the case of cesarean delivery. *Journal of Health Economics*, 18(4):473–490.

Gruber, J. and Owings, M. (1996). Physician financial incentives and cesarean section delivery. *The RAND Journal of Economics*, 27(1):99–123.

Hannan, R. L., Kagel, J. H., and Moser, D. V. (2002). Partial gift exchange in an experimental labor market: Impact of subject population differences, productivity differences, and effort requests on behavior*. *Journal of Labor Economics*, 20(4):923–951.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4):1009–1055.

Harter, J. K. and Arora, R. (2010). The impact of time spent working and job fit on well-being around the world. In Diener, E., Kahneman, D., and Helliwell, J., editors, *International differences in well-being*, pages 398–435. Oxford University Press, New York, NY.

Hay, J. and Leahy, M. J. (1982). Physician-induced demand: an empirical analysis of the consumer information gap. *Journal of Health Economics*, 1(3):231–244.

Hennig-Schmidt, H., Selten, R., and Wiesen, D. (2011). How payment systems affect physicians provision behaviouran experimental investigation. *Journal of Health Economics*, 30(4):637–646.

Hennig-Schmidt, H. and Wiesen, D. (2014). Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students. *Social Science & Medicine*, 108:156–165.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Huck, S., Lünser, G., Spitzer, F., and Tyran, J.-R. (2014). Medical Insurance and Free Choice of Physician Shape Patient Overtreatment. A Laboratory Experiment. Discussion Papers 14-19, University of Copenhagen. Department of Economics.

Huck, S., Lünser, G., and Tyran, J.-R. (2007). Pricing and trust. *UCL Working Paper*.

Huck, S., Lünser, G. K., and Tyran, J.-R. (2012). Competition fosters trust. *Games and Economic Behavior*, 76(1):195–209.

Hughes, D. and Yule, B. (1992). The effect of per-item fees on the behaviour of general practitioners. *Journal of Health Economics*, 11(4):413–437.

Iizuka, T. (2007). Experts' agency problems: evidence from the prescription drug market in Japan. *The RAND Journal of Economics*, 38(3):844–862.

Jürges, H. (2007). Health insurance status and physician-induced demand for medical services in Germany: New evidence from combined district and individual level data. *SSRN eLibrary*, SOEP Paper No. 8.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Kairies, N. and Krieger, M. (2013). How do non-monetary performance incentives for physicians affect the quality of medical care? : a laboratory experiment. Ruhr economic papers 2013-414.

Karsan, R. (2007). Calculating the cost of turnover. *Employment Relations Today*, 34(1):33–36.

Kerschbamer, R., Sutter, M., and Dulleck, U. (2009). The impact of distributional preferences on (experimental) markets for expert services. Working Paper 2009-28, Faculty of Economics and Statistics, University of Innsbruck.

Kerschbamer, R., Sutter, M., and Dulleck, U. (2015). How social preferences shape incentives on (experimental) markets for credence goods. *Forthcoming in Economic Journal*.

Keser, C., Montmarquette, C., Schmidt, M., and Schnitzler, C. (2014). Custom-made healthcare–an experimental investigation. *cege Discussion Papers*, (218).

Kessler, D. and McClellan, M. (1996). Do doctors practice defensive medicine? *The Quarterly Journal of Economics*, 111(2):353–390.

Kesternich, I., Schumacher, H., and Winter, J. (2014). Professional norms and physician behavior: homo oeconomicus or homo hippocraticus? *SFB/TR 15 Discussion Paper No. 456*, pages 1–29.

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 75(1):23–55.

Kühberger, A., Schulte-Mecklenbeck, M., and Perner, J. (1999). The effects of framing, reflection, probability, and payoff on risk preference in choice tasks. *Organizational Behavior and Human Decision Processes*, 78(3):204–231.

Levin, I. P., Schneider, S. L., and Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76(2):149–188.

Lichtenstein, S. and Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of experimental psychology*, 89(1):46–55.

Lichtenstein, S. and Slovic, P. (1973). Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of experimental psychology*, 101(1):16–20.

Mazar, N. and Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing*, 25(1):117–126.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2):311–329.

Oosterbeek, H., Sloof, R., and Van De Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, 7(2):171–188.

Pingle, M. (2010). Looking under the hood: Exploring assumptions and finding behavioral economics. *Journal of economic behavior & organization*, 73(1):73–76.

Pitchik, C. and Schotter, A. (1987). Honesty in a model of strategic information transmission. *The American Economic Review*, 77(5):1032–1036.

Potters, J. and van Winden, F. (2000). Professionals and students in a lobbying experiment: Professional rules of conduct and subject surrogacy. *Journal of Economic Behavior & Organization*, 43(4):499–522.

Rasch, A. and Waibel, C. (2015). What drives fraud in a credence goods market? evidence from a field study. Discussion Paper No. 180, Düsseldorf Institute for Competition Economics.

Schaffner, M. (2013). Programming for experimental economics: Introducing coral - a lightweight framework for experimental economic experiments. QuBE Working Papers 016, QUT Business School.

Schneider, H. S. (2012). Agency problems and reputation in expert services: Evidence from auto repair. *The Journal of Industrial Economics*, 60(3):406–433.

Scroggins, W. A. (2008). The relationship between employee fit perceptions, job performance, and retention: Implications of perceived fit. *Employee Responsibilities and Rights Journal*, 20(1):57–71.

Severinov, S. and Deneckere, R. (2006). Screening when some agents are nonstrategic: does a monopoly need to exclude? *The RAND Journal of Economics*, 37(4):816–840.

Siegel, S. and Harnett, D. (1964). Bargaining behavior: A comparison between mature industrial personnel and college students. *Operations Research*, 12(2):334–343.

Slonim, R., Wang, C., Garbarino, E., and Merrett, D. (2013). Opting-in: Participation bias in economic experiments. *Journal of Economic Behavior and Organization*, 90:43–70.

Slovic, P. and Lichtenstein, S. (1983). Preference reversals: A broader perspective. *The American Economic Review*, 73(4):596–605.

Smith, A. (1776). The wealth of nations. In Cannan, E., editor, *Reprint*. Modern Library, 1937, New York, NY.

Smith, V. L. (1962). An experimental study of competitive market behavior. *The Journal of Political Economy*, 70(2):111–137.

Sülzle, K. and Wambach, A. (2005). Insurance in a market for credence goods. *Journal of Risk and Insurance*, 72(1):159–176.

Taylor, C. R. (1995). The economics of breakdowns, checkups, and cures. *Journal of Political Economy*, 103(1):53–74.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458.

Tversky, A. and Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, 59(4):S251–S278.

Tversky, A., Slovic, P., and Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, 80(1):204–217.

Wilson, R. M., Runciman, W. B., Gibberd, R. W., Harrison, B. T., Newby, L., and Hamilton, J. D. (1995). The quality in Australian health care study. *The Medical journal of Australia*, 163(9):458–471.

Wolinsky, A. (1993). Competition in a market for informed experts' services. *The RAND Journal of Economics*, 24(3):380–398.

Zizzo, D. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.

Appendix A

Appendix to Chapter 3

A.1 Individual Participant Averages

Table A.1.1: Overview of Results - Neutral Framing

Averages per Period	**Baseline**	**Liability**	**Verifiability**
Undertreatment[a,b]	0.67	—	0.44
Overtreatment[a,c]	0.10	0.07	0.19
Overcharging[a,d]	0.81	0.70	—
Number of participants	32	32	32

[a] relative frequency.
[b] consumer needs t^h, but seller provides t^l.
[c] consumer needs t^l, but seller provides t^h.
[d] seller provides t^l, but charges p^h (with $p^l \leq p^h$ and consumer requiring t^l).

Table A.1.2: Overview of Results - Medical Framing

Averages per Period	**Baseline**	**Liability**	**Verifiability**
Undertreatment	0.33	—	0.50
Overtreatment	0.01	0.04	0.24
Overcharging	0.65	0.63	—
Number of participants	48	48	48

Note: See previous table's footnotes for a definition of the variables.

Table A.1.3: Overview Ranksum Test Results - Medical vs Neutral Framing

	Medical - Neutral Framing, Mean Diff. ($\|z - value^{e}\|$)		
Averages per Period	Baseline	Liability	Verifiability
Undertreatment[a,b]	-0.34 (4.38)***	—	0.06 (0.57)
Overtreatment[a,c]	-0.09 (3.56)***	-0.03 (0.65)	0.05 (1.04)
Overcharging[a,d]	-0.16 (2.14)**	-0.07 (0.76)	—

*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.
[a] relative frequency.
[b] consumer needs t^h, but seller provides t^l.
[c] consumer needs t^l, but seller provides t^h.
[d] seller provides t^l, but charges p^h (with $p^l \leq p^h$ and consumer requiring t^l).
[e] Wilcoxon Rank-Sum Test where Ho: $\text{variable}_i(\text{medical}==0) = \text{variable}_i(\text{medical}==1)$.

A.2 Difference between Baseline and Verifiability

Table A.2.1: Baseline vs Verifiability

	Baseline - Verifiability, Mean Diff. ($\lvert z - value^{g}\rvert$)	
Averages per Period	Neutral Framing	Medical Framing
Trade[a]	0.02 (0.75)	0.07 (2.78)***
Efficiency[b]	-0.06 (2.52)**	0.18 (10.96)***
Undertreatment[a,c]	0.24 (5.44)***	-0.17 (5.01)***
Overtreatment[a,d]	-0.08 (2.59)***	-0.24 (9.71)***
Overcharging[a,e]	———	
p^l with trade	-1.26 (8.37)***	-0.92 (9.48)***
p^l without trade	-1.42 (8.57)***	0.93 (7.06)***
p^h with trade	-1.08 (8.25)***	-0.38 (6.22)***
p^h without trade	-0.78 (6.17)***	-0.38 (4.02)***
Actually charged price	-0.18 (0.93)	0.67 (7.05)***
Profits sellers[f]	0.37 (6.43)***	0.55 (6.96)***
Profits consumers[f]	-0.51 (7.88)***	-0.05 (2.58)***

*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.
[a] relative frequency.
[b] calculated as (actual average profit - outside option) / (maximum possible average profit - outside option)
[c] consumer needs t^h, but seller provides t^l.
[d] consumer needs t^l, but seller provides t^h.
[e] seller provides t^l, but charges p^h (with $p^l \leq p^h$ and consumer requiring t^l).
[f] in experimental currency units, in medical framing sellers = doctors and consumers = patients.
[g] Wilcoxon Rank-Sum Test where Ho: variable$_i$(baseline==0) = variable$_i$(baseline==1).

A.3 Robustness Checks OLS Regressions

Table A.3.1: OLS Regression - Trade

Variables	Trade Baseline	Trade Liability	Trade Verifiability	Trade Overall
Medical	0.204***	0.003	0.070	0.090***
	(4.35)	(0.07)	(1.48)	(3.41)
Price Low	-0.012	-0.022***	-0.048***	-0.028***
	(-1.27)	(-2.64)	(-4.67)	(-5.08)
Price High	-0.148***	-0.155***	-0.125***	-0.133***
	(-14.57)	(-10.15)	(-8.33)	(-17.07)
Period 1-3	0.169***	-0.197***	0.197***	0.062***
	(5.20)	(-6.05)	(5.89)	(2.78)
Period 4-8	0.157***	-0.023	0.157***	0.099***
	(6.05)	(-0.83)	(4.88)	(5.64)
Liability				0.135***
				(4.16)
Verifiability				0.065*
				(1.93)
Constant	1.593***	2.077***	1.736***	1.651***
	(18.84)	(16.62)	(10.98)	(24.37)
F	75.983***	45.372***	31.816***	77.904***
$\Pr(F > f)$	0.000	0.000	0.000	0.000
R^2 Adjusted	0.172	0.258	0.153	0.178
N (Clusters)	80	80	80	240
N (Obs.)	1280	1280	1280	3840

Notes: Clustered standard errors, t-statistics in parentheses. The reference groups consist of NEUTRAL, PERIOD 9-16 and BASELINE. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table A.3.2: OLS Regression - Undertreatment

Variables	Undertreatment Baseline	Undertreatment Verifiability	Undertreatment Overall
Medical	-0.364***	0.077	-0.134**
	(-6.27)	(1.22)	(-2.50)
Price Low	-0.130***	0.127***	0.004
	(-8.34)	(7.25)	(0.23)
Price High	0.041**	-0.146***	-0.074***
	(2.50)	(-8.29)	(-4.54)
Period 1-3	-0.171***	0.020	-0.080**
	(-3.61)	(0.42)	(-2.12)
Period 4-8	-0.157***	-0.026	-0.088***
	(-4.09)	(-0.74)	(-2.97)
Verifiability			0.046
			(0.89)
Constant	1.044***	0.910***	1.139***
	(8.80)	(5.31)	(10.41)
F	66.879***	22.936***	12.822***
$\Pr(F > f)$	0.000	0.000	0.000
R^2Adjusted	0.299	0.238	0.071
N (Clusters)	80	80	160
N (Obs.)	666	648	1314

Notes: Clustered standard errors, t-statistics in parentheses. The reference groups consist of NEUTRAL, PERIOD 9-16 and BASELINE. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table A.3.3: OLS Regression - Overtreatment

Variables	Overtreatment Baseline	Overtreatment Liability	Overtreatment Verifiability	Overtreatment Overall
Medical	-0.093***	-0.015	0.040	-0.023
	(-3.41)	(-0.73)	(1.03)	(-1.14)
Price Low	0.016**	-0.008	-0.112***	-0.036***
	(2.16)	(-1.52)	(-8.26)	(-5.23)
Price High	-0.003	-0.013	0.102***	0.031***
	(-0.41)	(-1.21)	(7.16)	(3.51)
Period 1-3	0.115***	0.141***	0.022	0.130***
	(3.06)	(4.20)	(0.50)	(5.58)
Period 4-8	0.020	0.017	0.030	0.041***
	(1.36)	(1.59)	(0.88)	(2.98)
Liability				-0.008
				(-0.41)
Verifiability				0.199***
				(7.10)
Constant	0.031	0.168*	-0.037	-0.055
	(0.55)	(1.89)	(-0.28)	(-0.94)
F	3.207**	5.230***	22.771***	18.805***
$\Pr(F > f)$	0.011	0.000	0.000	0.000
R^2 Adjusted	0.102	0.080	0.229	0.136
N (Clusters)	80	80	80	240
N (Obs.)	614	624	632	1870

Notes: Clustered standard errors, t-statistics in parentheses. The reference groups consist of NEUTRAL, PERIOD 9-16 and BASELINE. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table A.3.4: OLS Regression - Overcharging

Variables	Overcharging Baseline	Overcharging Liability	Overcharging Overall
Medical	-0.119***	-0.108*	-0.110***
	(-2.86)	(-1.93)	(-3.13)
Price Low	-0.131***	-0.070***	-0.102***
	(-10.18)	(-5.74)	(-10.76)
Price High	0.054***	0.023*	0.041***
	(3.70)	(1.75)	(3.85)
Period 1-3	-0.085*	-0.394***	-0.206***
	(-1.84)	(-6.59)	(-5.41)
Period 4-8	-0.033	-0.086**	-0.051**
	(-1.13)	(-2.27)	(-2.13)
Liability			-0.075**
			(-2.03)
Constant	1.022***	0.975***	1.011***
	(9.89)	(9.55)	(12.74)
F	25.220***	35.974***	40.881***
$\Pr(F > f)$	0.000	0.000	0.000
R^2Adjusted	0.249	0.183	0.206
N (Clusters)	80	80	160
N (Obs.)	882	596	1478

Notes: Clustered standard errors, t-statistics in parentheses. The reference groups consist of NEUTRAL, PERIOD 9-16 and BASELINE. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Appendix B

Appendix to Chapter 4

B.1 Experimental Room Setup in Hamburg

Figure B.1.1: Room Setup DEGAM

B.2 Individual Participant Averages

Table B.2.1: Overview of Baseline Results Medical Professionals vs Students

	Neutral Framing		
Averages per Period	Medical Prof.	Students	Mean Diff. ($\lvert z - value^{e} \rvert$)
Undertreatment[a,b]	0.49	0.67	-0.18 (2.48)**
Overtreatment[a,c]	0.20	0.10	0.10 (2.63)***
Overcharging[a,d]	0.61	0.81	-0.20 (2.48)**
Number of participants	40	32	

*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.
[a] relative frequency.
[b] patient needs t^h, but seller provides t^l.
[c] patient needs t^l, but seller provides t^h.
[d] doctor provides t^l, but charges p^h (with $p^l \leq p^h$ and consumer requiring t^l).
[e] Wilcoxon Rank-Sum Test where Ho: variable$_i$($Medical_{prof}$ == 0) = variable$_i$($Medical_{prof}$ == 1).

B.3 Robustness Checks

B.3.1 Demographic Variables

Table B.3.1: Probit Regression - Including Demographic Variables

Variables	(5) Trade	(6) Undertreatment	(7) Overtreatment	(8) Overcharging
Medical Prof.	0.427**	-0.749***	0.603***	-1.172***
	(2.41)	(-3.77)	(3.10)	(-6.13)
	0.168	*-0.282*	*0.119*	*-0.259*
Price Low	-0.022	-0.388***	0.197***	-0.572***
	(-0.91)	(-8.21)	(6.08)	(-6.16)
	-0.009	*-0.150*	*0.041*	*-0.128*
Price High	-0.225***	0.156***	-0.006	0.405***
	(-6.98)	(3.47)	(-0.14)	(5.33)
	-0.089	*0.060*	*-0.001*	*0.091*
Period 1-3	0.373***	-0.495***	0.857***	-0.505***
	(3.80)	(-2.59)	(5.30)	(-2.91)
	0.143	*-0.194*	*0.231*	*-0.134*
Period 4-8	0.331***	-0.575***	0.146	-0.205*
	(4.22)	(-4.24)	(1.04)	(-1.78)
	0.128	*-0.223*	*0.032*	*-0.048*
Age	-0.007	0.010	0.005	0.009
	(-0.62)	(1.14)	(0.52)	(0.43)
	-0.003	*0.004*	*0.001*	*0.002*
Male	-0.149	-0.131	-0.020	-0.145
	(-0.89)	(-0.67)	(-0.12)	(-0.71)
	-0.059	*-0.051*	*-0.004*	*-0.033*
Constant	1.851***	1.258***	-2.607***	0.787
	(4.94)	(3.28)	(-5.66)	(1.47)
χ^2	69.032***	101.955***	144.998***	90.449***
$Prob > \chi^2$	0.000	0.000	0.000	0.000
(Pseudo) R^2	0.095	0.203	0.142	0.327
N (Clusters)	72	72	72	72
N (Obs.)	1152	550	602	824

Notes: Clustered standard errors, z-statistics in parentheses and marginal effects in italics. The reference groups consist of STUDENTS, PERIOD 9-16 and FEMALES. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

B.3.2 Results with QUT and IBK included

Table B.3.2: Probit Regression - IBK as Reference Group

Variables	(5) Trade	(6) Undertreatment	(7) Overtreatment	(8) Overcharging
Medical Prof.	0.350**	-0.549***	0.602***	-1.088***
	(2.21)	(-3.03)	(3.24)	(-6.08)
	0.137	*-0.214*	*0.138*	*-0.276*
QUT	0.177	-0.519***	0.380*	-0.642***
	(1.29)	(-2.60)	(1.92)	(-3.17)
	0.069	*-0.203*	*0.089*	*-0.160*
Price Low	-0.037**	-0.204***	0.150***	-0.533***
	(-2.17)	(-5.90)	(4.73)	(-8.58)
	-0.015	*-0.080*	*0.033*	*-0.116*
Price High	-0.179***	0.109***	-0.003	0.432***
	(-6.83)	(3.12)	(-0.11)	(7.43)
	-0.071	*0.043*	*-0.001*	*0.094*
Period 1-3	0.524***	-0.624***	0.894***	-0.474***
	(6.12)	(-4.31)	(6.90)	(-3.43)
	0.198	*-0.245*	*0.247*	*-0.122*
Period 4-8	0.386***	-0.504***	0.187	0.054
	(5.85)	(-4.91)	(1.52)	(0.49)
	0.150	*-0.197*	*0.042*	*0.012*
Constant	1.280***	0.861***	-2.311***	0.478*
	(5.89)	(3.02)	(-7.78)	(1.81)
χ^2	93.931***	104.296***	108.188***	137.022***
$Prob > \chi^2$	0.000	0.000	0.000	0.000
(Pseudo) R^2	0.084	0.115	0.124	0.320
N (Clusters)	104	104	104	104
N (Obs.)	1664	800	864	1182

Notes: Clustered standard errors, z-statistics in parentheses and marginal effects in italics. The reference groups consist of IBK STUDENTS and PERIOD 9-16. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

B.3.3 Results for Panel Regressions

Table B.3.3: Probit Random Effects Panel Regressions - Baseline

Variables	(1) Trade	(2) Undertreatment	(3) Overtreatment	(4) Overcharging
Medical Prof.	0.569***	-0.789***	0.747***	-1.377***
	(2.67)	(-3.35)	(3.26)	(-4.52)
	0.181	*-0.235*	*0.137*	*-0.251*
Price Low	-0.014	-0.379***	0.205***	-0.707***
	(-0.48)	(-7.60)	(5.23)	(-7.52)
	-0.004	*-0.113*	*0.037*	*-0.129*
Price High	-0.376***	0.160***	0.002	0.534***
	(-9.68)	(3.20)	(0.05)	(7.57)
	-0.120	*0.048*	*0.000*	*0.098*
Period 1-3	0.484***	-0.758***	0.981***	-0.744***
	(3.75)	(-3.27)	(5.30)	(-2.98)
	0.154	*-0.226*	*0.179*	*-0.136*
Period 4-8	0.455***	-0.735***	0.178	-0.457***
	(4.73)	(-4.53)	(1.11)	(-3.12)
	0.145	*-0.219*	*0.033*	*-0.084*
Constant	2.632***	1.541***	-2.867***	1.023***
	(8.07)	(3.79)	(-6.83)	(2.63)
χ^2	135.172***	90.001***	60.463***	115.999***
$Prob > \chi^2$	0.000	0.000	0.000	0.000
N (Clusters)	72	72	72	72
N (Obs.)	1152	550	602	824

Notes: Clustered standard errors, z-statistics in parentheses and marginal effects in italics. The reference groups consist of STUDENTS and PERIOD 9-16. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table B.3.4: Probit Random Effects Panel Regressions - Bootstrapped

Variables	(1) Trade	(2) Undertreatment	(3) Overtreatment	(4) Overcharging
Medical Prof.	0.569***	-0.789***	0.747***	-1.377***
	(5.44)	(-4.60)	(4.34)	(-5.11)
	0.181	*-0.235*	*0.137*	*-0.251*
Price Low	-0.014	-0.379***	0.205***	-0.707***
	(-0.52)	(-5.54)	(4.09)	(-6.86)
	-0.004	*-0.113*	*0.037*	*-0.129*
Price High	-0.376***	0.160***	0.002	0.534***
	(-9.30)	(2.94)	(0.05)	(5.96)
	-0.120	*0.048*	*0.000*	*0.098*
Period 1-3	0.484***	-0.758***	0.981***	-0.744***
	(2.95)	(-2.88)	(4.13)	(-2.87)
	0.154	*-0.226*	*0.179*	*-0.136*
Period 4-8	0.455***	-0.735***	0.178	-0.457***
	(4.24)	(-3.98)	(0.90)	(-2.63)
	0.145	*-0.219*	*0.033*	*-0.084*
Constant	2.632***	1.541***	-2.867***	1.023*
	(9.09)	(3.49)	(-6.13)	(1.84)
χ^2	120.245***	46.477***	36.625***	61.507***
$Prob > \chi^2$	0.000	0.000	0.000	0.000
N (Clusters)	72	72	72	72
N (Obs.)	1152	550	602	824

Notes: Bootstrapped standard errors (200 reps based on 72 clusters in number), z-statistics in parentheses and marginal effects in italics. The reference groups consist of STUDENTS and PERIOD 9-16. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table B.3.5: Random Effects Panel Regressions - Baseline

Variables	(1) Trade	(2) Undertreatment	(3) Overtreatment	(4) Overcharging
Medical Prof.	0.137**	-0.219***	0.129***	-0.272***
	(2.39)	(-3.66)	(3.52)	(-5.00)
Price Low	0.001	-0.101***	0.042***	-0.095***
	(0.11)	(-10.05)	(4.75)	(-7.86)
Price High	-0.090***	0.041***	0.008	0.068***
	(-11.09)	(3.23)	(0.95)	(9.53)
Period 1-3	0.114***	-0.191***	0.216***	-0.150***
	(3.37)	(-3.25)	(4.60)	(-3.35)
Period 4-8	0.119***	-0.188***	0.021	-0.062**
	(4.16)	(-4.56)	(0.73)	(-2.54)
Constant	1.107***	0.919***	-0.203***	0.761***
	(13.96)	(9.05)	(-2.85)	(12.22)
R^2 Overall	0.112	0.236	0.127	0.278
χ^2	157.065***	176.121***	94.302***	128.653***
$Prob > \chi^2$	0.000	0.000	0.000	0.000
N (Clusters)	72	72	72	72
N (Obs.)	1152	550	602	824

Notes: Clustered standard errors and z-statistics in parentheses. The reference groups consist of STUDENTS and PERIOD 9-16. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Appendix C

Appendix to Chapter 5

C.1 Ranksum Test Between Different Majors

Table C.1.1: Wilcoxon Ranksum Test Results

Majors	Health	Engineering	Law	Accounting
	Interact, Mean Diff. ($\lvert z-value \rvert$)			
Health	—			
Engineering	0.04 (1.05)	—		
Law	**0.15 (3.18)*****	**0.10 (2.45)****	—	
Accounting	0.07 (1.55)	0.02 (0.57)	**-0.08 (1.93)***	—
	Undertreatment, Mean Diff. ($\lvert z-value \rvert$)			
Health	—			
Engineering	**-0.10 (1.73)***	—		
Law	-0.04 (0.60)	0.06 (1.06)	—	
Accounting	0.00 (0.03)	**0.10 (1.92)***	0.04 (0.68)	—
	Overtreatment, Mean Diff. ($\lvert z-value \rvert$)			
Health	—			
Engineering	0.00 (0.06)	—		
Law	-0.02 (0.48)	-0.02 (0.62)	—	
Accounting	**0.05 (1.69)***	**0.05 (1.81)***	**0.07 (2.33)****	—
	Overcharging, Mean Diff. ($\lvert z-value \rvert$)			
Health	—			
Engineering	**-0.13 (2.49)****	—		
Law	-0.08 (1.29)	0.05 (1.13)	—	
Accounting	0.00 (0.07)	**0.12 (2.77)*****	0.07 (1.38)	—
	Accident Self Reporting, Mean Diff. ($\lvert z-value \rvert$)			
Health	—			
Engineering	-0.04 (0.74)	—		
Law	0.01 (0.24)	0.05 (1.02)	—	
Accounting	-0.05 (0.98)	-0.01 (0.26)	-0.06 (1.27)	—
	Tax Compliance Rate, Mean Diff.($\lvert z-value \rvert$)			
Health	—			
Engineering	0.00 0.18	—		
Law	0.05 0.79	0.05 (0.90)	—	
Accounting	-0.04 (1.39)	-0.04 (1.54)	**-0.09 (2.50)****	—
Participants	21	32	23	32

*, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

C.2 Descriptive Statistics

Table C.2.1: Summary Statistics for Health Participants

Variable	Mean	Std. Dev.	Min.	Max.	N
Trade	0.67	0.47	0	1	210
Undertreatment	0.29	0.46	0	1	114
Overtreatment	0.09	0.29	0	1	96
Overcharging	0.63	0.48	0	1	120
Payoff Doctor	2.47	2.28	-4	7	120
Payoff Patient	1.86	3.64	-9	8	90
Accident Self Reporting	0.30	0.46	0	1	142
Tax Compliance Rate	0.70	0.39	0	1.5	210

Table C.2.2: Summary Statistics for Engineering Participants

Variable	Mean	Std. Dev.	Min.	Max.	N
Trade	0.62	0.49	0	1	320
Undertreatment	0.39	0.49	0	1	156
Overtreatment	0.09	0.29	0	1	164
Overcharging	0.76	0.43	0	1	210
Payoff Doctor	3.29	1.97	-1	8	120
Payoff Patient	1.04	3.63	-8	7	200
Self Report	0.33	0.47	0	1	213
Tax Compliance Rate	0.70	0.39	0	1.5	320

Table C.2.3: Summary Statistics for Law Participants

Variable	Mean	Std. Dev.	Min.	Max.	N
Trade	0.52	0.50	0	1	230
Undertreatment	0.33	0.47	0	1	107
Overtreatment	0.11	0.32	0	1	123
Overcharging	0.71	0.46	0	1	144
Payoff Doctor	2.52	2.40	-4	9	110
Payoff Patient	1.85	3.28	-11	9	120
Accident Self Reporting	0.28	0.45	0	1	152
Tax Compliance Rate	0.65	0.41	0	1.2	230

Table C.2.4: Summary Statistics for Accounting Participants

Variable	Mean	Std. Dev.	Min.	Max.	N
Trade	0.60	0.49	0	1	320
Undertreatment	0.29	0.45	0	1	153
Overtreatment	0.04	0.20	0	1	167
Overcharging	0.64	0.48	0	1	204
Payoff Doctor	2.69	1.95	0	9	190
Payoff Patient	1.84	2.86	-8	8	130
Accident Self Reporting	0.35	0.48	0	1	223
Tax Compliance Rate	0.74	0.37	0	1.5	320

Appendix D

Appendix to Experiments

D.1 Experimental Instructions

D.1.1 Experimental Instructions Baseline Neutral

Thank you very much for participating in this experiment. Please do not to talk to any other participant until the end of the experiment. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds**, each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.

There are 2 roles in this experiment: **Player A** and **Player B**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

Player A always interacts with Player B. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by Player A, Decision 2 is made by Player B. At the start of the round Player B is randomly assigned a type (Type I or Type II).

Sequence of Decisions in a Round (in short)

1. Player A chooses a price for Action I and a price for Action II.
2. Player B is informed about the prices chosen by Player A. Player B then decides whether they want to interact with Player A.
3. Player A (but **not** Player B) is informed about the type of Player B. Then Player A is asked to choose either Action I or Action II.
4. Player A charges one of the prices specified in Decision 1. The price charged **does not** necessarily have to be the price for the action actually chosen by Player A in Decision 3, it can also be the price of the other action.
5. The results of this round, based on the decisions of Player A and Player B, will be announced.

Detailed Description of the Decisions and their Consequences in Terms of Payoffs

Decision 1

Player A will later choose between two actions, Action I and Action II.
For **Action I** Player A incurs costs of **2 points** (=experimental currency).
For **Action II** Player A incurs costs of **6 points**.
Player A can charge prices for these actions from Player B. In **Decision 1** Player A has to set the **prices** (in points) **for both actions**. Both prices have to be positive integers between 1 and 11, i.e., only 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 are valid prices. Also, the price for Action I must not exceed the price for Action II.

Decision 2

Player B is informed about the prices set by Player A for the two actions in Decision 1. Then Player B decides whether they want to interact with Player

A.

If Player B decides **not to interact**, the round **ends** and both players receive a **payoff of 1.6 points for this round**.

If Player B decides to interact, Player A can choose an action in Decision 3 and charge a price for this action in Decision 4. Player B is **not** able to observe which action is chosen by Player A.

Decision 3

At the beginning of the round Player B was randomly assigned a type. **Player B** can be one of two types: **Type I** or **Type II**.

Player B is **with a 50**% of **Type I** and **with a 50**% of **Type II**. Imagine that a coin is tossed for each Player B in each round. If the coin comes up "heads", Player B is of Type I, if it comes up as "tails", Player B is of Type II.

Player A is informed about the type of Player B *before* they enter Decision 3. Then Player A chooses an action for Player B, either Action I or Action II.

An **action** is **sufficient** under the following conditions:

- Player B is of Type I and Player A either chooses Action I or Action II.
- Player B is of Type II and Player A chooses Action II..

An action is **insufficient**, if Player B is of Type II and Player A chooses Action I.

Player B receives **10 points**, if Player A did choose a **sufficient action**. **Player B** receives **0 points** if Player A did choose an **insufficient action**.

Player B will never be informed whether they are of Type I or Type II or which action Player A has chosen.

Decision 4

Player A **charges** Player B one of the two **prices** specified in Decision 1 for the two actions. This price does **not** have to be the price of the action actually chosen in Decision 3, but could also be the price of the other action.

Payoffs

If Player B chooses not to interact with Player A (*decision* "***Don't interact***" by Player B) both players get **1.6 points** for the round.

Otherwise (decision "***Interact***" by Player B) the payoffs are as follows:
Player A receives the **price** (denoted in points) charged in Decision 4 **minus** the **costs** for the action chosen in Decision 3.

Player B's payoff depends on whether the action chosen by Player A in Decision 3 was sufficient.

- If the action chosen by Player A was sufficient, Player B receives 10 points minus the price charged in Decision 4. .

- If the action was insufficient, Player B has to pay the price charged in Decision 4.

At the beginning of the experiment you receive an initial **endowment of 10 points**. You are able to cover losses that might occur in some rounds with this initial endowment. Furthermore, losses can also be compensated by gains in other rounds. If your total payoff is negative at the end of the experiment, you will have to pay this amount to the supervisor of the experiment. By participating in this experiment you agree to this condition. Please note that it is **always** possible to avoid losses in this experiment.

To calculate your total earnings the initial endowment and the profits from all rounds are added up. This sum is then converted into cash using the following exchange rate:

1 point = 50 Cent
(i.e. 2 points = 1 Dollar)

D.1.2 Experimental Instructions Baseline Medical

Thank you very much for participating in this experiment. Please do not to talk to any other participant until the end of the experiment. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds**, each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.

There are 2 roles in this experiment: **Doctors** and **Patients**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

A doctor always interacts with a patient. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by doctors; Decision 2 is made by patients. In each round, a patient has either a minor or a severe illness. Without knowing the gravity of their illness, a patient has to decide if they want to be examined by a doctor or not. If a patient seeks an examination, a doctor will then perform a diagnosis by which they can determine the gravity of the illness for certain. Following the diagnosis a doctor will then administer a normal or a special intravenous drip to a patient.

Sequence of Decisions in a Round (in short)

1. The doctor chooses a price for the normal intravenous drip and a price for the special intravenous drip.

2. The patient is informed about the prices chosen by the doctor. The patient then decides whether they want to be examined by the doctor.

3. If the doctor performs a diagnosis, then the doctor (however **not** the patient) learns for certain if the patient has the minor or severe illness. The doctor then selects the normal or the special intravenous drip.

4. The doctor charges from the patients one of the prices specified in Decision 1. The price charged **does not** necessarily have to be the price for the intravenous drip actually chosen by the doctor in Decision 3, it can also be the price of the other intravenous drip.

5. The results of this round, based on the decisions of the doctor and the patient, will be announced.

Detailed Description of the Decisions and their Consequences in Terms of Payoffs

Decision 1
The **doctor** will later choose between two infusions, the normal and special intravenous drip.
For the **normal intravenous drip** doctors incur costs of **2 points** (=experimental currency).
For the **special intravenous drip** doctors incur costs of **6 points**.
Doctors can charge prices for these infusions from patients. In **Decision 1** the doctor has to set the **price** (in points) **for both infusions**. Both prices have to be positive integers between 1 and 11, i.e., only 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 are valid prices. Also, the price for the normal intravenous drip must not exceed the price for the special intravenous drip.

Decision 2

The **patient** is informed about the prices set by the doctor for the two infusions in Decision 1. Then the patient decides whether they want to be examined by the doctor.

If the patient decides **not to be examined**, the round **ends** and both participants receive a **payoff of 1.6 points for this round**.

If the patient decides to be examined, the doctor can choose an intravenous drip in Decision 3 and charge a price for this intravenous drip in Decision 4. The patient is **not** able to observe which intravenous drip is chosen by the doctor.

Decision 3

At the beginning of the round a patient was randomly assigned an illness. A **patient** can have one of two illnesses: **the minor illness** or **the severe illness**.

Patients have **with a 50% chance the minor illness** and **with a 50% chance the severe illness**. Imagine that a coin is tossed for each patient in each round. If the coin comes up "heads", the patient has the minor illness, if it comes up as "tails", the patient has the severe illness.

The doctor is informed about the illness of the patient *before* they enter Decision 3. Then the doctor chooses an intravenous drip for the patient, either the normal intravenous drip or the special intravenous drip.

An **intravenous drip** is **sufficient** under the following conditions:

- The patient has the minor illness and the doctor either administers the normal or the special intravenous drip.
- The patient has the severe illness and the doctor administers the special intravenous drip.

An intravenous drip is **insufficient**, if the patient has the severe illness and the doctor administered the normal intravenous drip.

The **patient** receives **10 points**, if the doctor administered a **sufficient in-**

travenous drip. The patient receives **0 points** if the doctor administered an **insufficient intravenous drip**.

The **patient** will never be informed whether they have the minor illness or the severe illness and which intravenous drip the doctor administered.

Decision 4

The Doctor **charges** the patient one of the two **prices** specified in Decision 1 for the infusions. This price does **not** have to be the price of the intravenous drip actually chosen in Decision 3, but could also be the price of the other intravenous drip.

Payoffs

If the patient chooses not to be examined by the doctor (*decision* "***No examination***" by the patient) both players get **1.6 points** for the round.

Otherwise (decision "***Examination***" by the patient) the payoffs are as follows: The **doctor** receives the **price** (denoted in points) charged in Decision 4 **minus** the **costs** for the intravenous drip administered in Decision 3.

The **patients** payoff depends on whether the intravenous drip administered by the doctor in Decision 3 was sufficient.

- If the intravenous drip administered by the doctor was sufficient, the patient receives 10 points minus the price charged in Decision 4.
- If the intravenous drip administered was insufficient, the patient has to pay the price charged in Decision 4.

At the beginning of the experiment you receive an initial **endowment of 10 points**. You are able to cover losses that might occur in some rounds with this initial endowment. Furthermore, losses can also be compensated by gains in other rounds. If your total payoff is negative at the end of the experiment, you will have to pay this amount to the supervisor of the experiment. By participating in this experiment you agree to this condition. Please note that it is **always** possible to avoid losses in this experiment.

To calculate your total earnings the initial endowment and the profits from all rounds are added up. This sum is then converted into cash using the following exchange rate:

1 point = 50 Cent

(i.e. 2 points = 1 Dollar)

D.1.3 Experimental Instructions Liability Neutral

Thank you very much for participating in this experiment. Please do not to talk to any other participant. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds**, each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.

There are 2 roles in this experiment: **Player A** and **Player B**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

Player A always interacts with Player B. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by Player A, Decision 2 is made by Player B. At the start of the round Player B is randomly assigned a type (Type I or Type II).

Sequence of Decisions in a Round (in short)

1. Player A chooses a price for Action I and a price for Action II.

2. Player B is informed about the prices chosen by Player A. Player B then decides whether they want to interact with Player A.

3. Player A (but **not** Player B) is informed about the type of Player B. Then Player A is asked to choose either Action I or Action II.

4. Player A charges one of the prices specified in Decision 1. The price charged **does not** necessarily have to be the price for the action actually chosen by Player A in Decision 3, it can also be the price of the other action.

5. The results of this round, based on the decisions of Player A and Player B, will be announced.

Detailed Description of the Decisions and their Consequences in Terms of Payoffs

Decision 1

Player A will later choose between two actions, Action I and Action II.
For **Action I** Player A incurs costs of **2 points** (=experimental currency).
For **Action II** Player A incurs costs of **6 points**.
Player A can charge prices for these actions from Player B. In **Decision 1** Player A has to set the **prices** (in points) **for both actions**. Both prices have to be positive integers between 1 and 11, i.e., only 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 are valid prices. Also, the price for Action I must not exceed the price for Action II.

Decision 2

Player B is informed about the prices set by Player A for the two actions in Decision 1. Then Player B decides whether they want to interact with Player A.
If Player B decides **not to interact**, the round **ends** and both players receive a **payoff of 1.6 points for this round**.

If Player B decides to interact, Player A can choose an action in Decision 3 and charge a price for this action in Decision 4. Player B is **not** able to observe which action is chosen by Player A.

Decision 3

At the beginning of the round Player B was randomly assigned a type. **Player B** can be one of two types: **Type I** or **Type II**.

Player B is **with a 50%** of **Type I** and **with a 50%** of **Type II**. Imagine that a coin is tossed for each Player B in each round. If the coin comes up "heads", Player B is of Type I, if it comes up as "tails", Player B is of Type II.

Player A is informed about the type of Player B *before* they enter Decision 3. Then Player A chooses an action for Player B, either Action I or Action II. Player A can choose between the following possibilities:

- If Player B is of Type I, Player A can either choose Action I or Action II.
- If Player B is of Type II, Player A can only choose Action II.

Player B will never be informed whether they are of Type I or Type II or which action Player A has chosen.

Decision 4

Player A **charges** Player B one of the two **prices** specified in Decision 1 for the two actions. This price does **not** have to be the price of the action actually chosen in Decision 3, but could also be the price of the other action.

Payoffs

If Player B chooses not to interact with Player A (*decision* "***Don't interact***" by Player B) both players get **1.6 points** for the round.

Otherwise (decision "***Interact***" by Player B) the payoffs are as follows:

Player A receives the **price** (denoted in points) charged in Decision 4 **minus** the **costs** for the action chosen in Decision 3.

Player B receives **10 points minus** the **price** charged in Decision 4.

At the beginning of the experiment you receive an initial **endowment of 10 points**. You are able to cover losses that might occur in some rounds with this initial endowment. Furthermore, losses can also be compensated by gains in other rounds. If your total payoff is negative at the end of the experiment, you will have to pay this amount to the supervisor of the experiment. By participating in this experiment you agree to this condition. Please note that it is **always** possible to avoid losses in this experiment.

To calculate your total earnings the initial endowment and the profits from all rounds are added up. This sum is then converted into cash using the following exchange rate:

1 point = 50 Cent

(i.e. 2 points = 1 Dollar)

D.1.4 Experimental Instructions Liability Medical

Thank you very much for participating in this experiment. Please do not to talk to any other participant. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds**, each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.

There are 2 roles in this experiment: **Doctors** and **Patients**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

A doctor always interacts with a patient. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by doctors; Decision 2 is made by patients. In each round, a patient has either a minor or a severe illness. Without knowing the gravity of their illness, a patient has to decide if they want to be examined by a doctor or not. If a patient seeks an examination, a doctor will then perform a diagnosis by which they can determine the gravity of the illness for certain. Following the diagnosis a doctor will then administer a normal or a special intravenous drip to a patient.

Sequence of Decisions in a Round (in short)

1. The doctor chooses a price for the normal intravenous drip and a price for the special intravenous drip.

2. The patient is informed about the prices chosen by the doctor. The patient then decides whether they want to be examined by the doctor.

3. If the doctor performs a diagnosis, then the doctor (however **not** the patient) learns for certain if the patient has the minor or severe illness. The doctor then selects the normal or the special intravenous drip.

4. The doctor charges from the patients one of the prices specified in Decision 1. The price charged **does not** necessarily have to be the price for the intravenous drip actually chosen by the doctor in Decision 3, it can also be the price of the other intravenous drip.

5. The results of this round, based on the decisions of the doctor and the patient, will be announced.

Detailed Description of the Decisions and their Consequences in Terms of Payoffs

Decision 1

The **doctor** will later choose between two infusions, the normal and special intravenous drip.

For the **normal intravenous drip** doctors incur costs of **2 points** (=experimental currency).

For the **special intravenous drip** doctors incur costs of **6 points**.

Doctors can charge prices for these infusions from patients. In **Decision 1** the doctor has to set the **price** (in points) **for both infusions**. Both prices have to be positive integers between 1 and 11, i.e., only 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 are valid prices. Also, the price for the normal intravenous drip must not exceed the price for the special intravenous drip.

Decision 2

The **patient** is informed about the prices set by the doctor for the two infusions in Decision 1. Then the patient decides whether they want to be examined by the doctor.

If the patient decides **not to be examined**, the round **ends** and both participants receive a **payoff of 1.6 points for this round**.

If the patient decides to be examined, the doctor can choose an intravenous drip in Decision 3 and charge a price for this intravenous drip in Decision 4. The patient is **not** able to observe which intravenous drip is chosen by the doctor.

Decision 3

At the beginning of the round a patient was randomly assigned an illness. A **patient** can have one of two illnesses: **the minor illness** or **the severe illness**.

Patients have **with a 50% chance the minor illness** and **with a 50% chance the severe illness**. Imagine that a coin is tossed for each patient in each round. If the coin comes up "heads", the patient has the minor illness, if it comes up as "tails", the patient has the severe illness.

The doctor is informed about the illness of the patient *before* they enter Decision 3. Then the doctor chooses an intravenous drip for the patient, either the normal intravenous drip or the special intravenous drip.

A doctor can choose between the following possibilities:

- If the patient has the minor illness, the doctor can either administer the normal or the special intravenous drip.

- If the patient has the severe illness, the doctor can only administer the special intravenous drip.

The **patient** will never be informed whether they have the minor illness or the severe illness and which intravenous drip the doctor administered.

Decision 4

The Doctor **charges** the patient one of the two **prices** specified in Decision 1 for the infusions. This price does **not** have to be the price of the intravenous drip actually chosen in Decision 3, but could also be the price of the other intravenous drip.

Payoffs

If the patient chooses not to be examined by the doctor (*decision* "***No examination***" by the patient) both players get **1.6 points** for the round.

Otherwise (decision "***Examination***" by the patient) the payoffs are as follows: The **doctor** receives the **price** (denoted in points) charged in Decision 4 **minus** the **costs** for the intravenous drip administered in Decision 3.

The **patient** receives **10 points minus** the **price** charged in Decision 4.

At the beginning of the experiment you receive an initial **endowment of 10 points**. You are able to cover losses that might occur in some rounds with this initial endowment. Furthermore, losses can also be compensated by gains in other rounds. If your total payoff is negative at the end of the experiment, you will have to pay this amount to the supervisor of the experiment. By participating in this experiment you agree to this condition. Please note that it is **always** possible to avoid losses in this experiment.

To calculate your total earnings the initial endowment and the profits from all rounds are added up. This sum is then converted into cash using the following exchange rate:

1 point = 50 Cent

(i.e. 2 points = 1 Dollar)

D.1.5 Experimental Instructions Verifiability Neutral

Thank you very much for participating in this experiment. Please do not to talk to any other participant until the end of the experiment. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds**, each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.

There are 2 roles in this experiment: **Player A** and **Player B**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

Player A always interacts with Player B. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 3 decisions, which are made sequentially. Decisions 1 and 3 are made by Player A, Decision 2 is made by Player B. At the start of the round Player B is randomly assigned a type (Type I or Type II).

Sequence of Decisions in a Round (in short)

1. Player A chooses a price for Action I and a price for Action II.

2. Player B is informed about the prices chosen by Player A. Player B then decides whether they want to interact with Player A.

3. Player A (but **not** Player B) is informed about the type of Player B. Then Player A is asked to choose either Action I or Action II. Player B has to pay the previously chosen price for the respective Action.

4. The results of this round, based on the decisions of Player A and Player B, will be announced.

Detailed Description of the Decisions and their Consequences in Terms of Payoffs

Decision 1

Player A will later choose between two actions, Action I and Action II.
For **Action I** Player A incurs costs of **2 points** (=experimental currency).
For **Action II** Player A incurs costs of **6 points**.
Player A can charge prices for these actions from Player B. In **Decision 1** Player A has to set the **prices** (in points) **for both actions**. Both prices have to be positive integers between 1 and 11, i.e., only 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 are valid prices. Also, the price for Action I must not exceed the price for Action II.

Decision 2

Player B is informed about the prices set by Player A for the two actions in Decision 1. Then Player B decides whether they want to interact with Player A.
If Player B decides **not to interact**, the round **ends** and both players receive a **payoff of 1.6 points for this round**.
If Player B decides to interact, Player A can choose an action in Decision 3 and charge the specified price for this action.

Decision 3

At the beginning of the round Player B was randomly assigned a type. **Player B** can be one of two types: **Type I** or **Type II**.

Player B is **with a 50**% of **Type I** and **with a 50**% of **Type II**. Imagine that a coin is tossed for each Player B in each round. If the coin comes up “heads”, Player B is of Type I, if it comes up as “tails”, Player B is of Type II.
Player A is informed about the type of Player B *before* they enter Decision 3. Then Player A chooses an action for Player B, either Action I or Action II.
An **action** is **sufficient** under the following conditions:

- Player B is of Type I and Player A either chooses Action I or Action II.
- Player B is of Type II and Player A chooses Action II..

An action is **insufficient**, if Player B is of Type II and Player A chooses Action I.

Player B receives **10 points**, if Player A did choose a **sufficient action**. **Player B** receives **0 points** if Player A did choose an **insufficient action**.

Player B will never be informed whether they are of Type I or Type II or which action Player A has chosen.

Player A **charges** from Player B the **price** specified in Decision 1 for the actions chosen in Decision 3.

Payoffs

If Player B chooses not to interact with Player A (*decision* “***Don’t interact***” by Player B) both players get **1.6 points** for the round.

Otherwise (decision “***Interact***” by Player B) the payoffs are as follows:
Player A receives the **price** (denoted in points) specified in Decision 1 (for the action chosen in Decision 3) **minus** the **costs** for the action chosen in Decision 3.

Player B’s payoff depends on whether the action chosen by Player A in Decision

3 was sufficient.

- If the action chosen by Player A was sufficient, Player B receives 10 points minus the price specified in Decision 1 for the action chosen in Decision 3.
- If the action chosen was insufficient, Player B has to pay the price for the action chosen in Decision 3.

At the beginning of the experiment you receive an initial **endowment of 10 points**. You are able to cover losses that might occur in some rounds with this initial endowment. Furthermore, losses can also be compensated by gains in other rounds. If your total payoff is negative at the end of the experiment, you will have to pay this amount to the supervisor of the experiment. By participating in this experiment you agree to this condition. Please note that it is **always** possible to avoid losses in this experiment.

To calculate your total earnings the initial endowment and the profits from all rounds are added up. This sum is then converted into cash using the following exchange rate:

1 point = 50 Cent
(i.e. 2 points = 1 Dollar)

D.1.6 Experimental Instructions Verifiability Medical

Thank you very much for participating in this experiment. Please do not to talk to any other participant until the end of the experiment. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds**, each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.

There are 2 roles in this experiment: **Doctors** and **Patients**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

A doctor always interacts with a patient. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 3 decisions, which are made sequentially. Decisions 1 and 3 are made by doctors; Decision 2 is made by patients. In each round, a patient has either a minor or a severe illness. Without knowing the gravity of their illness, a patient has to decide if they want to be examined by a doctor or not. If a patient seeks an examination, a doctor will then perform a diagnosis by which they can determine the gravity of the illness for certain. Following the diagnosis a doctor will then administer a normal or a special intravenous drip to a patient.

Sequence of Decisions in a Round (in short)

1. The doctor chooses a price for the normal intravenous drip and a price for the special intravenous drip.

2. The patient is informed about the prices chosen by the doctor. The patient then decides whether they want to be examined by the doctor.

3. If the doctor performs a diagnosis, then the doctor (however **not** the patient) learns for certain if the patient has the minor or severe illness. The doctor then selects the normal or the special intravenous drip. The patient has to pay the previously chosen price for the respective intravenous drip.

4. The results of this round, based on the decisions of the doctor and the patient, will be announced.

Detailed Description of the Decisions and their Consequences in Terms of Payoffs

Decision 1

The **doctor** will later choose between two infusions, the normal and special intravenous drip.

For the **normal intravenous drip** doctors incur costs of **2 points** (=experimental currency).

For the **special intravenous drip** doctors incur costs of **6 points**.

Doctors can charge prices for these infusions from patients. In **Decision 1** the doctor has to set the **price** (in points) **for both infusions**. Both prices have to be positive integers between 1 and 11, i.e., only 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 are valid prices. Also, the price for the normal intravenous drip must not exceed the price for the special intravenous drip.

Decision 2

The **patient** is informed about the prices set by the doctor for the two infusions in Decision 1. Then the patient decides whether they want to be examined by the doctor.

If the patient decides **not to be examined**, the round **ends** and both participants receive a **payoff of 1.6 points for this round**.

If the patient decides to be examined, the doctor can choose an intravenous drip in Decision 3 and charge the specified price for this intravenous drip.

Decision 3

At the beginning of the round a patient was randomly assigned an illness. A **patient** can have one of two illnesses: **the minor illness** or **the severe illness**. Patients have **with a 50% chance the minor illness** and **with a 50% chance**

the severe illness. Imagine that a coin is tossed for each patient in each round. If the coin comes up "heads", the patient has the minor illness, if it comes up as "tails", the patient has the severe illness.

The doctor is informed about the illness of the patient *before* they enter Decision 3. Then the doctor chooses an intravenous drip for the patient, either the normal intravenous drip or the special intravenous drip.

An **intravenous drip** is **sufficient** under the following conditions:

- The patient has the minor illness and the doctor either administers the normal or the special intravenous drip.
- The patient has the severe illness and the doctor administers the special intravenous drip.

An intravenous drip is **insufficient**, if the patient has the severe illness and the doctor administered the normal intravenous drip.

The **patient** receives **10 points**, if the doctor administered a **sufficient intravenous drip**. The patient receives **0 points** if the doctor administered an **insufficient intravenous drip**.

The **patient** will never be informed whether they have the minor illness or the severe illness and which intravenous drip the doctor administered.

The doctor **charges** from the patient the **price** specified in Decision 1 for the intravenous drip chosen in Decision 3.

Payoffs

If the patient chooses not to be examined by the doctor (*decision* "***No examination***" by the patient) both players get **1.6 points** for the round.

Otherwise (decision "***Examination***" by the patient) the payoffs are as follows: The **doctor** receives the **price** (denoted in points) specified in Decision 1 (for the intravenous drip chosen in Decision 3) **minus** the **costs** for the intravenous drip administered in Decision 3.

The **patients** payoff depends on whether the intravenous drip administered by the doctor in Decision 3 was sufficient.

- If the intravenous drip administered by the doctor was sufficient, the patient receives 10 points minus the price specified in Decision 1 for the intravenous drip chosen in Decision 3.
- If the intravenous drip administered was insufficient, the patient has to pay the price for the intravenous drip chosen in Decision 3.

At the beginning of the experiment you receive an initial **endowment of 10 points**. You are able to cover losses that might occur in some rounds with this initial endowment. Furthermore, losses can also be compensated by gains in other rounds. If your total payoff is negative at the end of the experiment, you will have to pay this amount to the supervisor of the experiment. By participating in this experiment you agree to this condition. Please note that it is **always** possible to avoid losses in this experiment.

To calculate your total earnings the initial endowment and the profits from all rounds are added up. This sum is then converted into cash using the following exchange rate:

1 point = 50 Cent

(i.e. 2 points = 1 Dollar)

D.1.7 Experimental Instructions Triple Experiment

Task 1

This task consists of **10 Rounds** each of which has the same sequence of decisions. The sequence of decisions is explained in detail below. There are 2 roles in this task: **Doctors** and **Patients**. At the beginning of the task you will be randomly assigned to one role and remain in that role throughout the task. On the first screen of the task you will see which role you have been assigned to.

A doctor always interacts with a patient. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by doctors; Decision 2 is made by patients. In each round, a patient has either a minor or a severe illness. Without knowing the gravity of their illness, a patient has to decide if they want to be examined by a doctor or not. If a patient seeks an examination, a doctor will then perform a diagnosis by which they can determine the gravity of the illness for certain. Following the diagnosis a doctor will then administer a normal or a special intravenous drip to a patient.

Description of the Decisions and their Consequences in Terms of Payoffs

Decision 1 (Doctors)

The **doctor** will later choose between two infusions, the normal and special intravenous drip.

For the **normal intravenous drip** doctors incur costs of **100 lab$**.

For the **special intravenous drip** doctors incur costs of **300 lab$**.

Doctors can charge prices for these infusions from patients. In **Decision 1** the doctor has to set the **price** (in lab$) **for both infusions**. Both prices have to be positive multiples of 50 between 50 and 550, i.e., only 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 or 550 are valid prices. Also, the price for the normal intravenous drip must not exceed the price for the special intravenous drip.

Decision 2 (Patients)

The **patient** is informed about the prices set by the doctor for the two infusions in Decision 1. Then the patient decides whether they want to be examined by the doctor.

If the patient decides **not to be examined**, the round ends and both participants receive a **payoff of 80 lab$ for this round**.

If the patient decides to be examined, the doctor can choose an intravenous drip in Decision 3 and charge a price for this intravenous drip in Decision 4. The patient is **not** able to observe which intravenous drip is chosen by the doctor.

Decision 3 (Doctors)

At the beginning of the round a patient was randomly assigned an illness. A **patient** can have one of two illnesses: **the minor illness** or **the severe illness**. Patients have **with a 50% chance** the **minor illness** and **with a 50% chance** the **severe illness**. Imagine that a coin is tossed for each patient in each round. If the coin comes up "heads", the patient has the minor illness, if it comes up as "tails", the patient has the severe illness. **The doctor is informed** about the illness of the patient when they enter Decision 3. Then the doctor chooses an intravenous drip for the patient, either the normal intravenous drip or the special intravenous drip.

An **intravenous drip** is **sufficient** under the following conditions:

- The patient has the minor illness and the doctor either administers the normal or the special intravenous drip.

- The patient has the severe illness and the doctor administers the special intravenous drip.

An **intravenous drip** is **insufficient**, if the patient has the severe illness and the doctor administered the normal intravenous drip.

The **patient** receives **500 lab$** minus the price charged in Decision 4, if the doctor administered a **sufficient intravenous drip**. The **patient** receives **0 lab$** minus the price charged in Decision 4, if the doctor administered an **insufficient intravenous drip**.

The **patient** will **never** be informed whether they have the minor illness or the severe illness and which intravenous drip the doctor administered.

Decision 4 (Doctors)
The Doctor **charges** the patient one of the two **prices** specified in Decision 1 for the infusions. This price does **not** have to be the price of the intravenous drip actually chosen in Decision 3, but could also be the price of the other intravenous drip.

Task 2

This task consists of **10 Rounds** each of which has the same sequence of decisions. In this task you are responsible for making a production decision as the owner of a chemical production company. When you produce chemicals, there is a chance that a chemical spill will occur. Your production decision directly affects the probability of a chemical spill. Reducing the probability of a chemical spill is costly and will reduce your production earnings. Similarly increasing the probability of a chemical spill will increase your production earnings.

Description of the Decisions and their Consequences in Terms of Payoffs

You will receive your production earnings **regardless** of whether a chemical spill occurs. However there is **a chance** that you **will be inspected**. If you are inspected and if a chemical spill occurred, then you will incur a fine. The probability that you will be **inspected is 50%** and if a chemical **spill has occurred** then you have to **pay a fine of 75lab$**.

The options below show you the relationship between the probability of a chemical spill and your production earnings. For example if you choose the probability of a chemical spill to be 50%, then your production earnings are equal to 107 lab$. If a chemical spill occurs and you are inspected, you would have to pay a fine of 75 lab$. Your earnings in this case would be 107 - 75 = 32 lab$. This occurs 20% of the time (0.5*0.4 = 20%). In the remaining 80% of the cases you would earn 107 lab$.

Please choose one of the following options:

- ☐ 100% accident probability and production earnings of 138 lab$
- ☐ 80% accident probability and production earnings of 132 lab$
- ☐ 60% accident probability and production earnings of 123 lab$
- ☐ 40% accident probability and production earnings of 107 lab$
- ☐ 20% accident probability and production earnings of 76 lab$

You will pay a fine of 75 lab$ if a chemical spill occurs and if you are inspected. Otherwise you will not pay a fine and you will earn 107 lab$.

Whether or not you have a chemical spill will be **determined by the computer** in accordance with your chosen **accident probability** and is **independent** across rounds. This means that whether or not you have a chemical spill this period is **not affected** by what happened **last period**. Similarly whether you are **inspected or not** is also **determined by the computer** and is **not affected** by previous inspection outcomes.

In addition, you will be asked to **fill in a report** about whether an accident has occurred or not. If you **report that you had an accident**, you will **pay** a self-reporting **fine of 60 lab$**. If you **report** that you have **not had an accident**, you will **be inspected with 50% probability** and **fined 75 lab$** if an **accident did occur**.

Task 3

This task consists of **10 rounds** each of which has the same sequence of decisions. In this task you are responsible for making tax declaration decisions. Each round represents one year and proceeds in three stages. All the information to make an informed decision will be provided on the screen.

For this task you will be **grouped together with 4** other participants. Please know that the tax revenue is redistributed to the members of your group.

Declare your income

First Stage

You have been **assigned an income of 100 lab$**. This will represent your **taxable income** for each of the following rounds (years). There is a **universal tax rate of 30%**, you will only be taxed on the amount of income you declare each round.

Tax calculation and audit

Second Stage

In the second stage each year, the taxes payable on your declared income will be deducted from your income. Your net income accumulates as 'wealth'. In this stage, there is also **a 10% chance** that you will be **selected for audit**, and your declared income will be checked against your real income in this year. If there is a discrepancy between your stated and your real income, a **fine of 1.5 times the amount of undeclared income** will be deducted from your income this year on top of all taxes owed.

Tax Income redistribution

Third Stage

Finally, every group member will receive a transfer, which is the sum of the tax contributions of all members of the group multiplied by 1.6 and divided by the number of group members (4). So if the amount of taxes (i.e. the sum of tax payments of all members of the group) is 30 lab$, each group member receives a transfer payment of 12 lab$ (= 30 * 1.6 / 4).

D.2 Print Screens

D.2.1 Print Screens Baseline Neutral

D.2.1.1 Player A & Player B

Experimental Instructions

Thank you very much for participating in this experiment. Please do not to talk to any other participant. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds** each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.
There are 2 roles in this experiment: **Player A** and **Player B**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

Player A always interacts with Player B. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by Player A, Decision 2 is made by Player B. At the start of the round Player B is randomly assigned a type (Type I or Type II).

Sequence of Decisions in a Round (in short)

1. Player A chooses a price for Action I and a price for Action II.
2. Player B is informed about the prices chosen by Player A. Player B then decides whether they want to interact with Player A.
3. Player A (but **not** Player B) is informed about the type of Player B. Then Player A is asked to choose either Action I or Action II.
4. Player A charges one of the prices specified in Decision 1. The price charged **does not** necessarily have to be the price for the action actually chosen by Player A in Decision 3, it can also be the price of the other action.
5. The results of this round, based on the decisions of Player A and Player B, will be announced.

Figure D.2.1: Screenshot Instruction Baseline Neutral

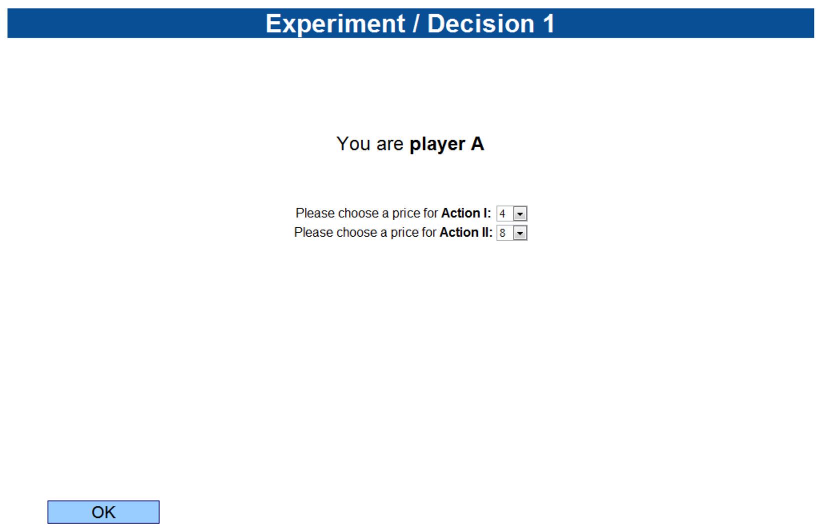


Figure D.2.2: Screenshot Player A / Decision 1

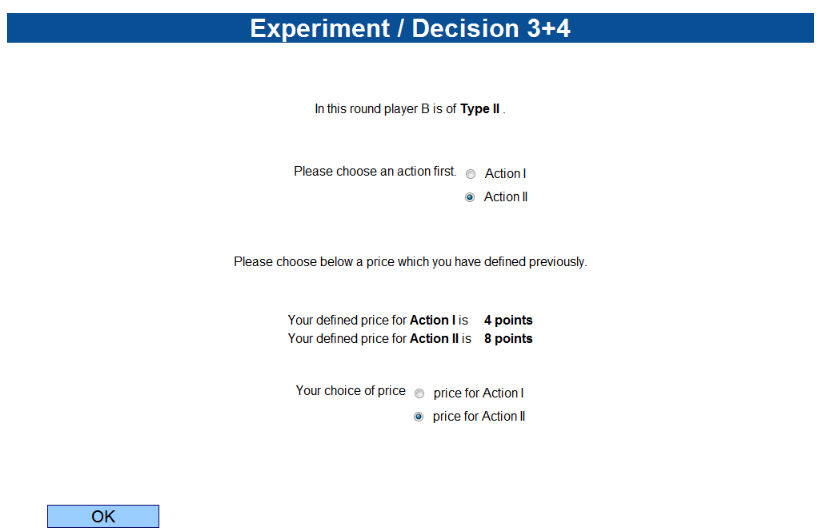


Figure D.2.3: Screenshot Player A / Decision 3 + 4

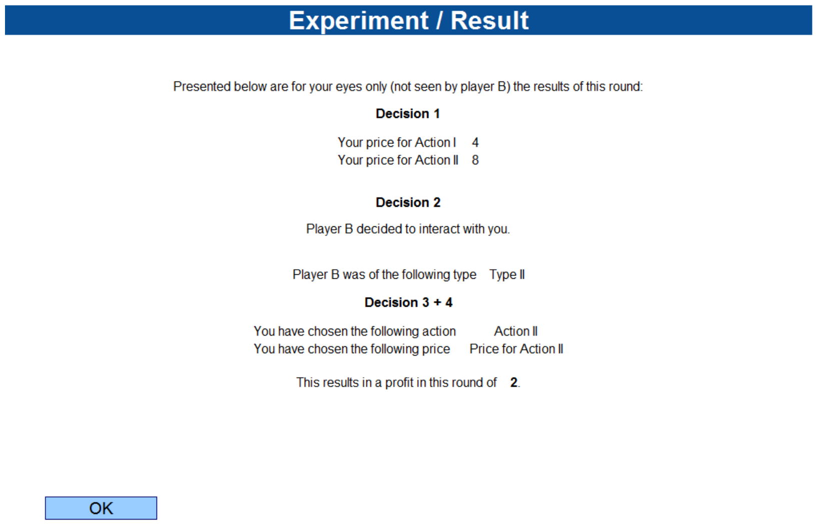


Figure D.2.4: Screenshot Player A / Result

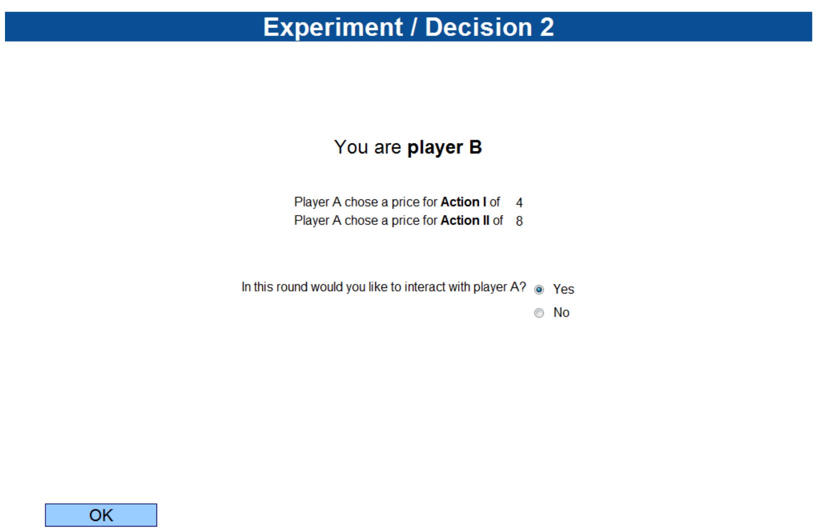


Figure D.2.5: Screenshot Player B / Decision 2

Experiment / Result

Presented below are the results of this round:

Decision 1

Player A has chosen the following prices at Decision 1

Price for Action I 4
Price for Action II 8

Decision 2

In Decision 2 you have decided to interact with player A.

Decision 3 + 4

Player A's action at Decision 3 was: sufficient

At Decision 4 Player A has chosen the following price: Price for Action II

This results in a profit in this round of **2**.

OK

Figure D.2.6: Screenshot Player B / Result

Experiment / Finished

The experiment is now finished. Please wait until we have your payments ready.

Thank you for your participation.

Figure D.2.7: Screenshot Player A & Player B / Final Screen

D.2.1.2 Print Screen When Outside Option was Chosen

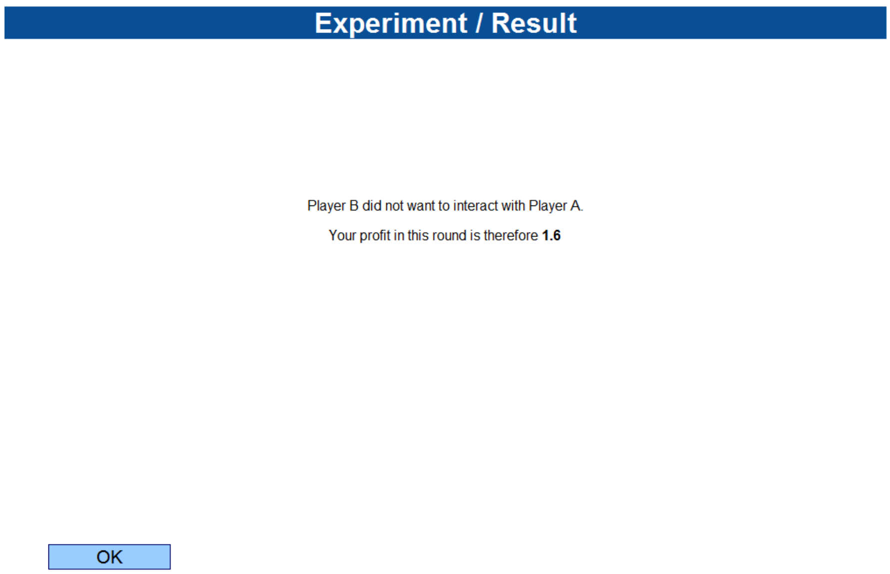
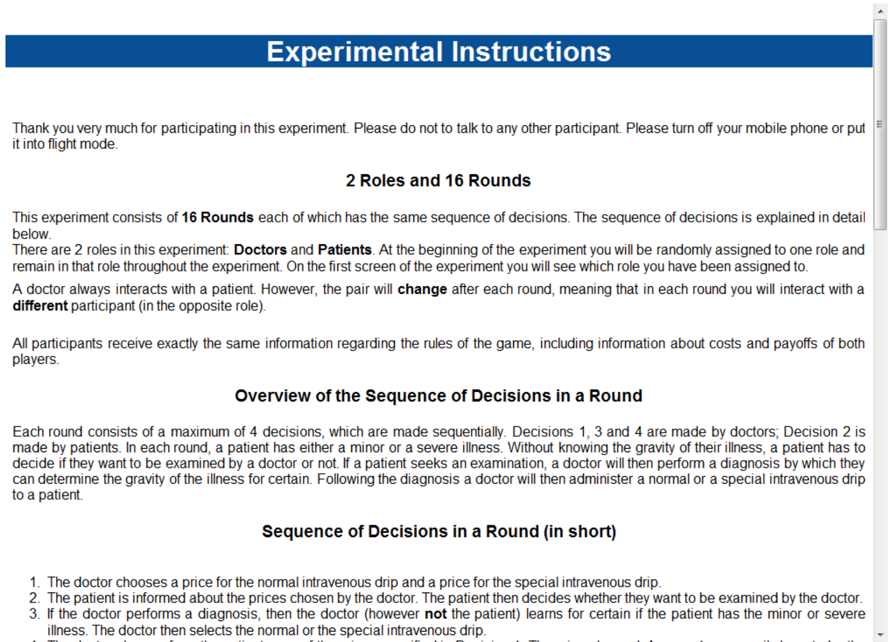


Figure D.2.8: Screenshot Player A & Player B / Outside Option

D.2.2 Print Screens Baseline Medical

D.2.2.1 Doctor & Patient


Experimental Instructions

Thank you very much for participating in this experiment. Please do not to talk to any other participant. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds** each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.
There are 2 roles in this experiment: **Doctors** and **Patients**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

A doctor always interacts with a patient. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by doctors; Decision 2 is made by patients. In each round, a patient has either a minor or a severe illness. Without knowing the gravity of their illness, a patient has to decide if they want to be examined by a doctor or not. If a patient seeks an examination, a doctor will then perform a diagnosis by which they can determine the gravity of the illness for certain. Following the diagnosis a doctor will then administer a normal or a special intravenous drip to a patient.

Sequence of Decisions in a Round (in short)

1. The doctor chooses a price for the normal intravenous drip and a price for the special intravenous drip.
2. The patient is informed about the prices chosen by the doctor. The patient then decides whether they want to be examined by the doctor.
3. If the doctor performs a diagnosis, then the doctor (however **not** the patient) learns for certain if the patient has the minor or severe illness. The doctor then selects the normal or the special intravenous drip.

Figure D.2.9: Screenshot Instruction Baseline Medical

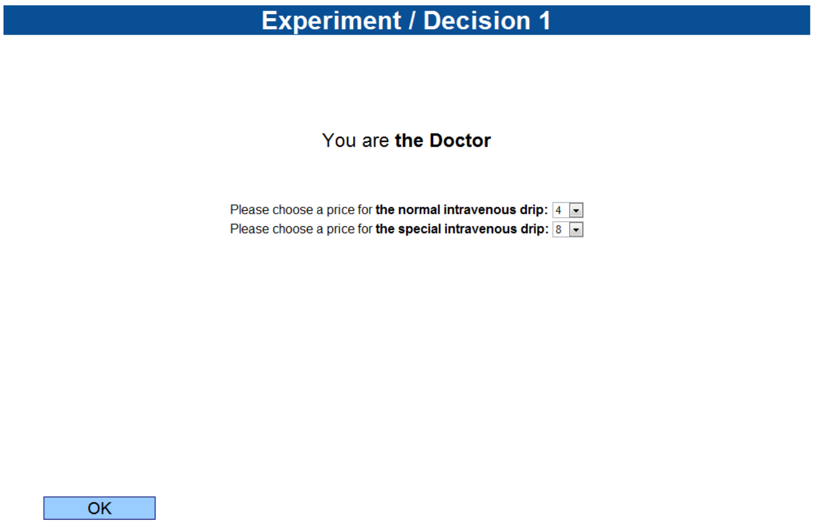


Figure D.2.10: Screenshot Doctor / Decision 1

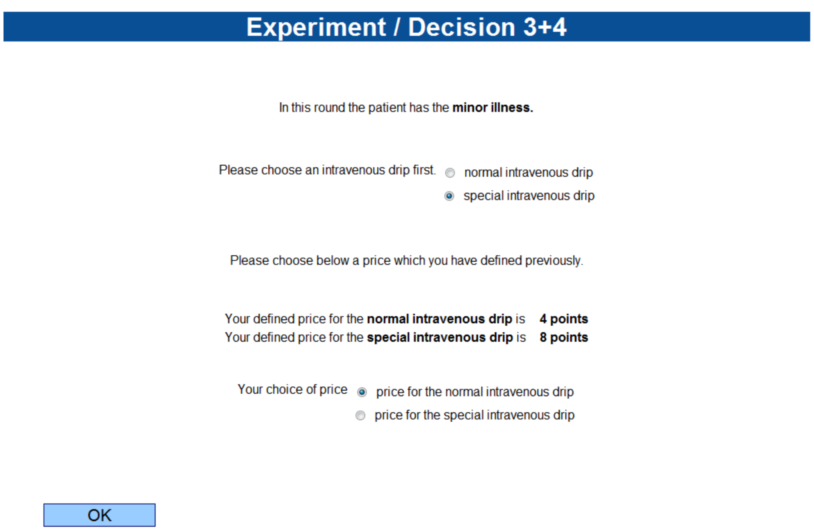


Figure D.2.11: Screenshot Doctor / Decision 3 + 4

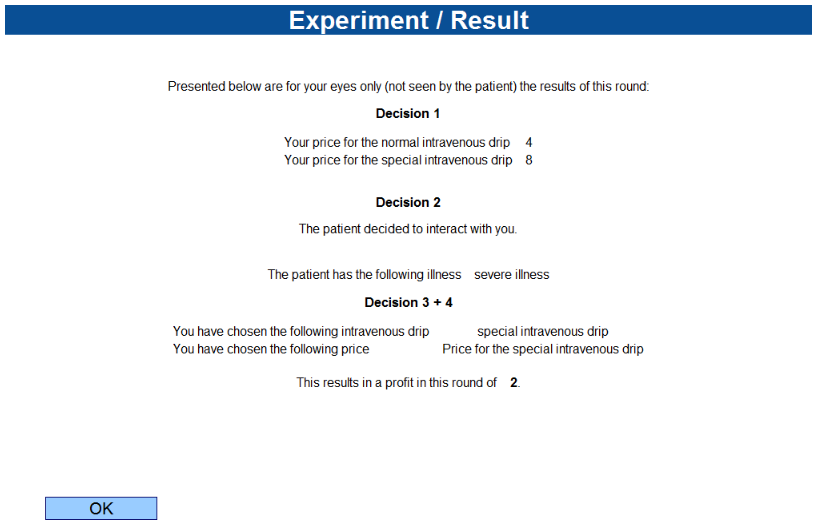


Figure D.2.12: Screenshot Doctor / Result

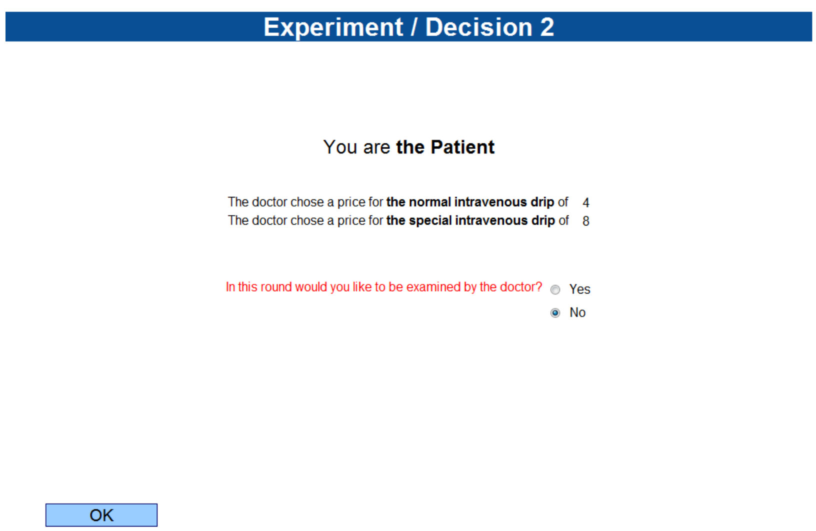


Figure D.2.13: Screenshot Patient / Decision 2

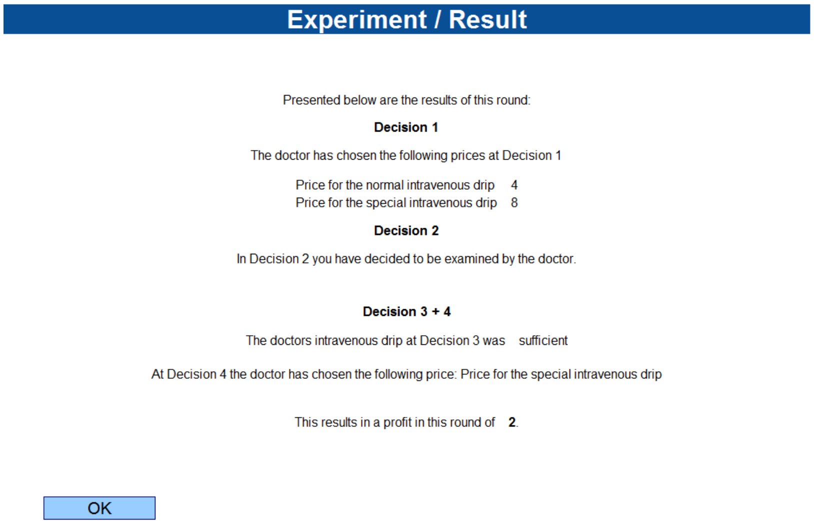


Figure D.2.14: Screenshot Patient / Result

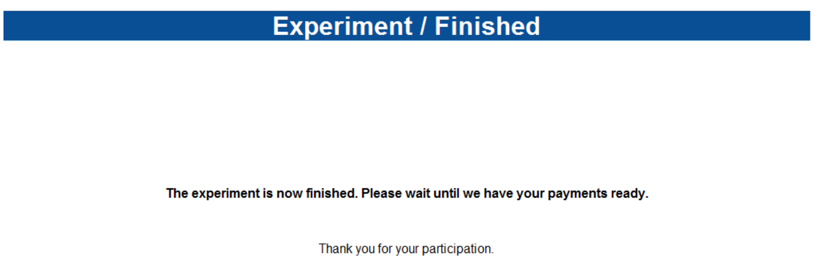


Figure D.2.15: Screenshot Doctor & Patient / Final Screen

D.2.2.2 Print Screen When Outside Option was Chosen

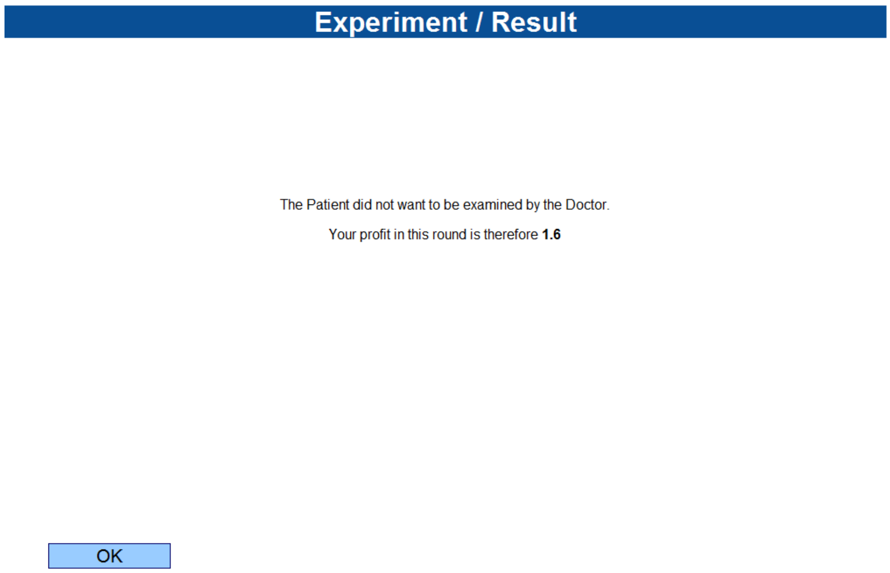
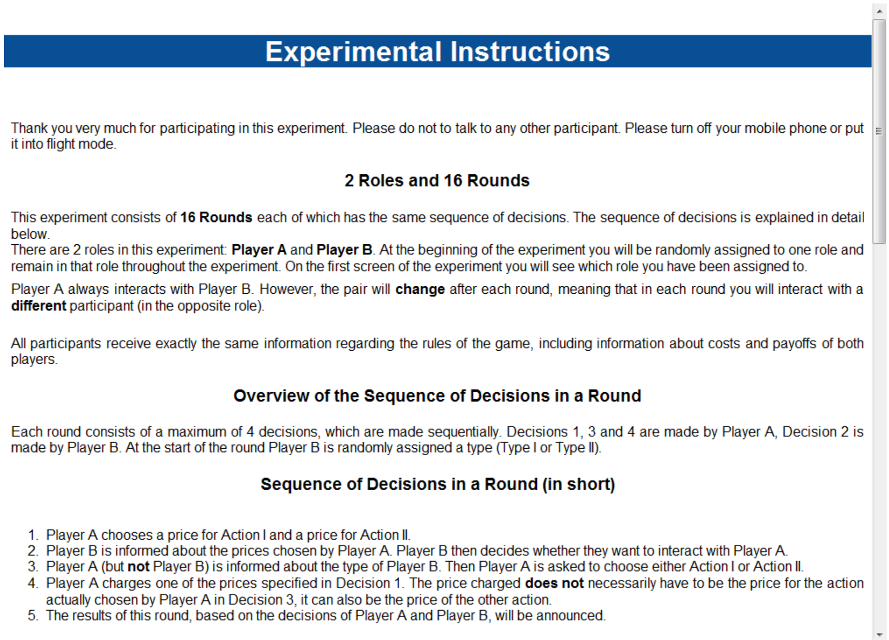


Figure D.2.16: Screenshot Doctor & Patient / Outside Option

D.2.3 Print Screens Liability Neutral

D.2.3.1 Player A & Player B



Experimental Instructions

Thank you very much for participating in this experiment. Please do not to talk to any other participant. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds** each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.
There are 2 roles in this experiment: **Player A** and **Player B**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

Player A always interacts with Player B. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by Player A, Decision 2 is made by Player B. At the start of the round Player B is randomly assigned a type (Type I or Type II).

Sequence of Decisions in a Round (in short)

1. Player A chooses a price for Action I and a price for Action II.
2. Player B is informed about the prices chosen by Player A. Player B then decides whether they want to interact with Player A.
3. Player A (but **not** Player B) is informed about the type of Player B. Then Player A is asked to choose either Action I or Action II.
4. Player A charges one of the prices specified in Decision 1. The price charged **does not** necessarily have to be the price for the action actually chosen by Player A in Decision 3, it can also be the price of the other action.
5. The results of this round, based on the decisions of Player A and Player B, will be announced.

Figure D.2.17: Screenshot Instruction Liability Neutral

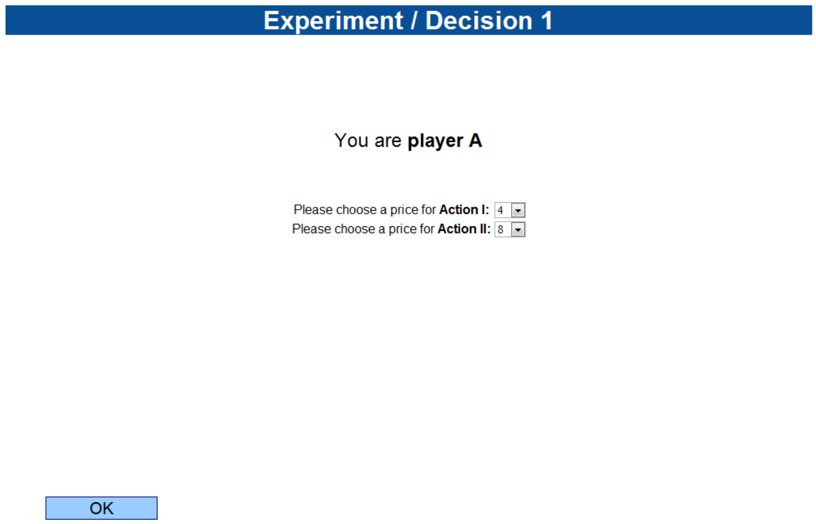


Figure D.2.18: Screenshot Player A / Decision 1

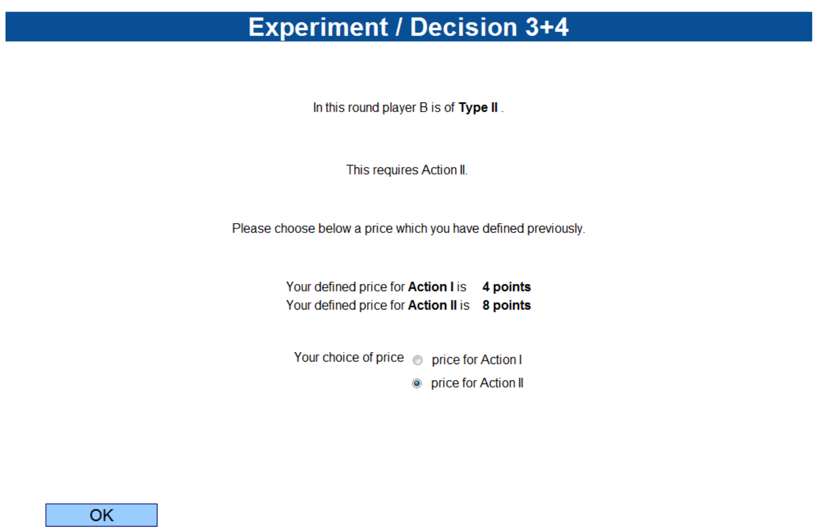


Figure D.2.19: Screenshot Player A / Decision 3 + 4

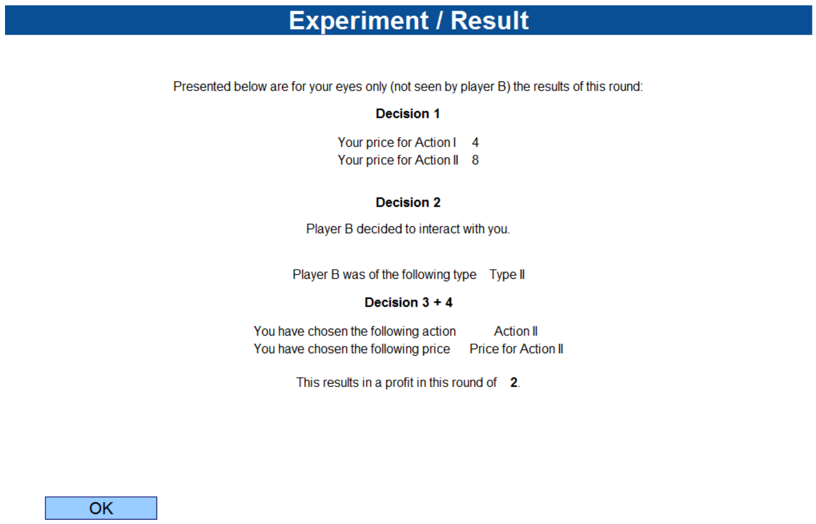


Figure D.2.20: Screenshot Player A / Result

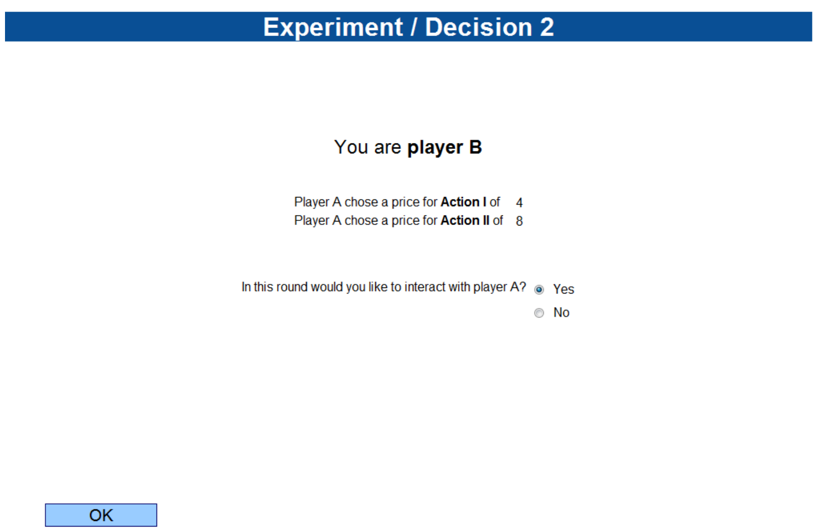


Figure D.2.21: Screenshot Player B / Decision 2

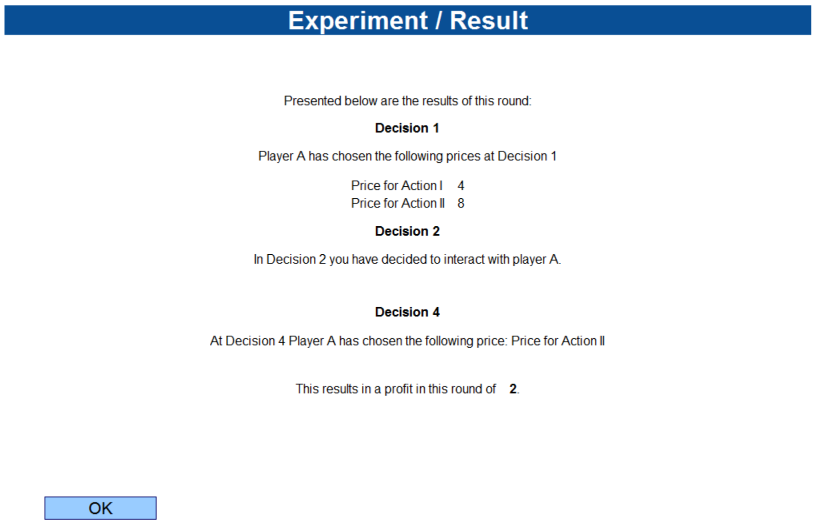


Figure D.2.22: Screenshot Player B / Result

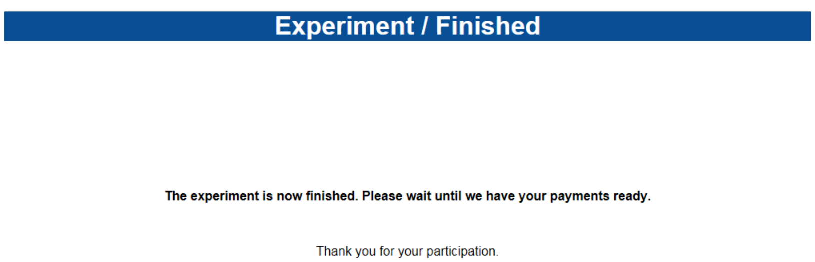


Figure D.2.23: Screenshot Player A & Player B / Final Screen

D.2.3.2 Print Screen When Outside Option was Chosen

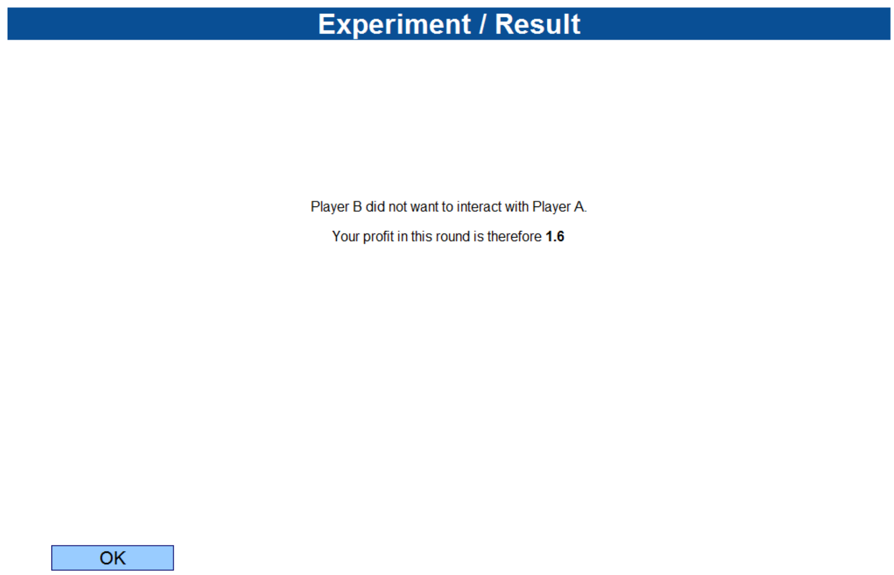
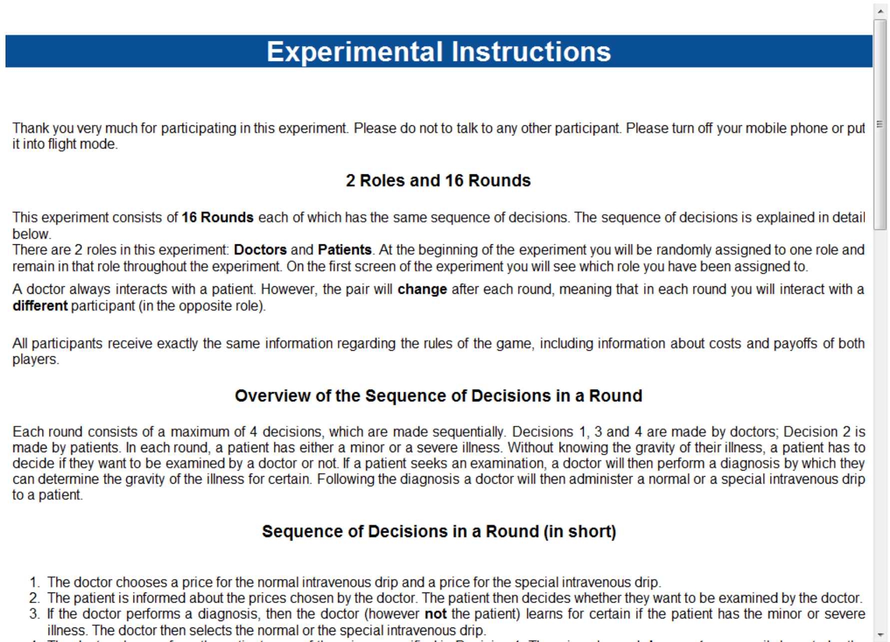


Figure D.2.24: Screenshot Player A & Player B / Outside Option

D.2.4 Print Screens Liability Medical

D.2.4.1 Doctor & Patient



Experimental Instructions

Thank you very much for participating in this experiment. Please do not to talk to any other participant. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds** each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.
There are 2 roles in this experiment: **Doctors** and **Patients**. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.

A doctor always interacts with a patient. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by doctors; Decision 2 is made by patients. In each round, a patient has either a minor or a severe illness. Without knowing the gravity of their illness, a patient has to decide if they want to be examined by a doctor or not. If a patient seeks an examination, a doctor will then perform a diagnosis by which they can determine the gravity of the illness for certain. Following the diagnosis a doctor will then administer a normal or a special intravenous drip to a patient.

Sequence of Decisions in a Round (in short)

1. The doctor chooses a price for the normal intravenous drip and a price for the special intravenous drip.
2. The patient is informed about the prices chosen by the doctor. The patient then decides whether they want to be examined by the doctor.
3. If the doctor performs a diagnosis, then the doctor (however **not** the patient) learns for certain if the patient has the minor or severe illness. The doctor then selects the normal or the special intravenous drip.

Figure D.2.25: Screenshot Instruction Liability Medical

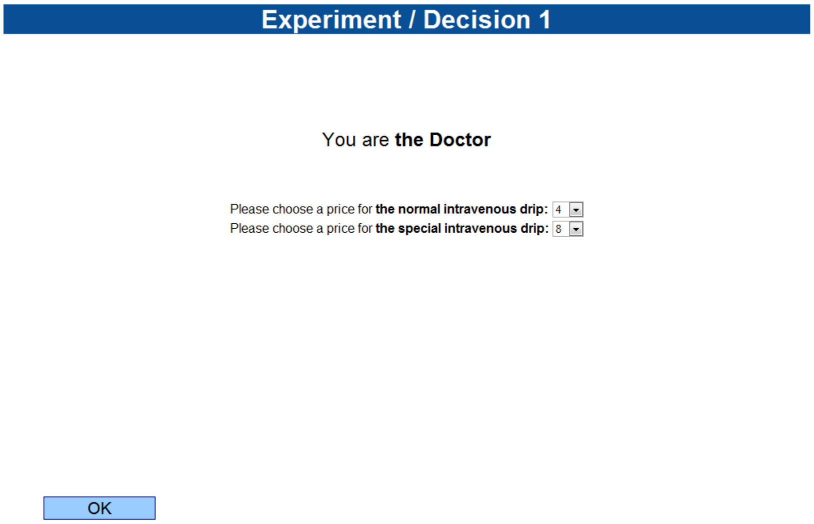


Figure D.2.26: Screenshot Doctor / Decision 1

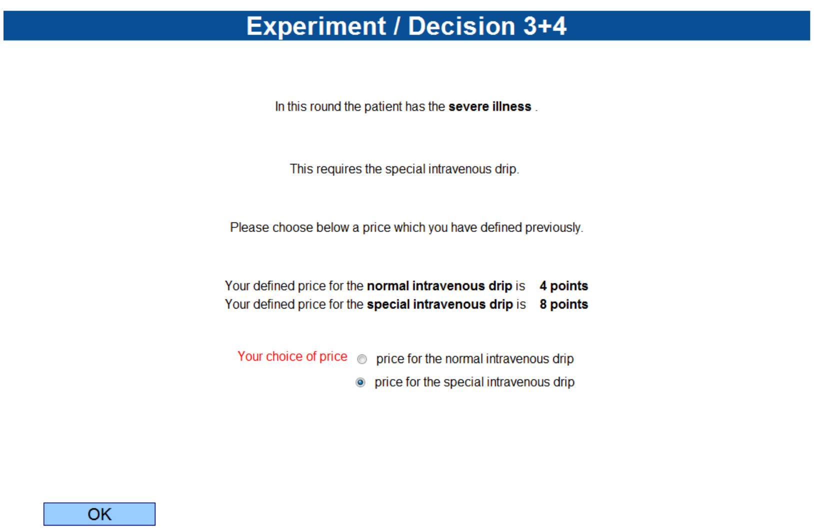


Figure D.2.27: Screenshot Doctor / Decision 3 + 4

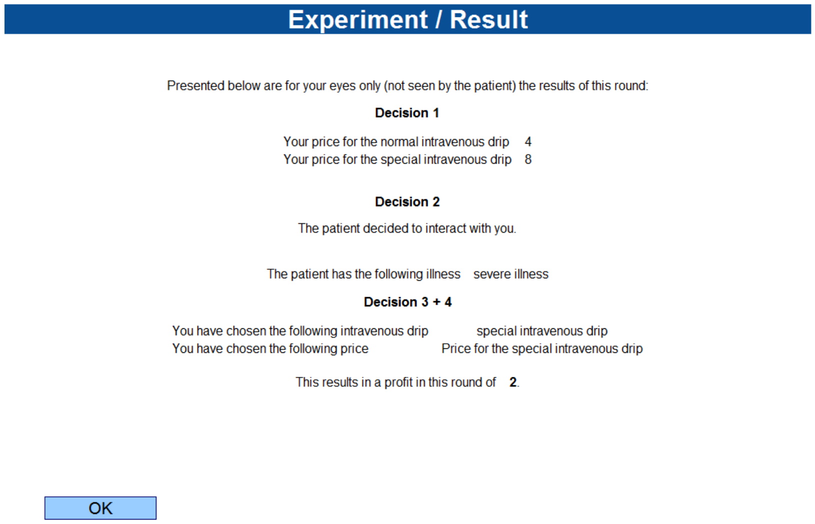


Figure D.2.28: Screenshot Doctor / Result

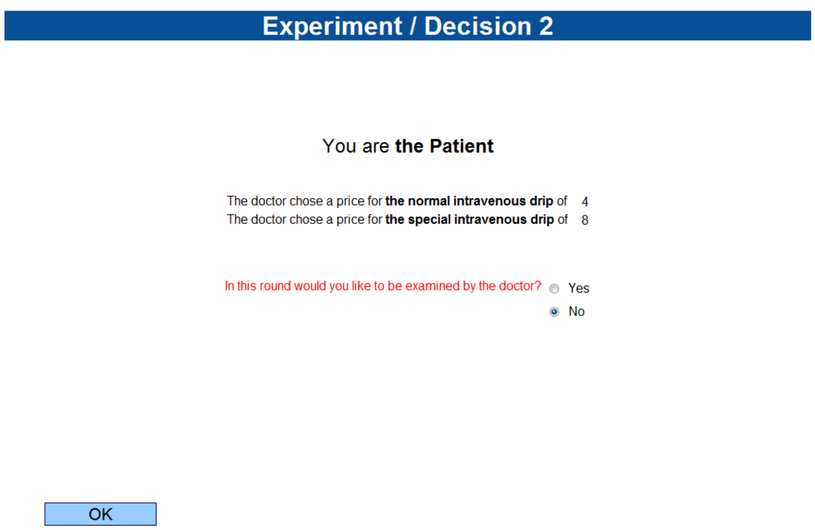


Figure D.2.29: Screenshot Patient / Decision 2

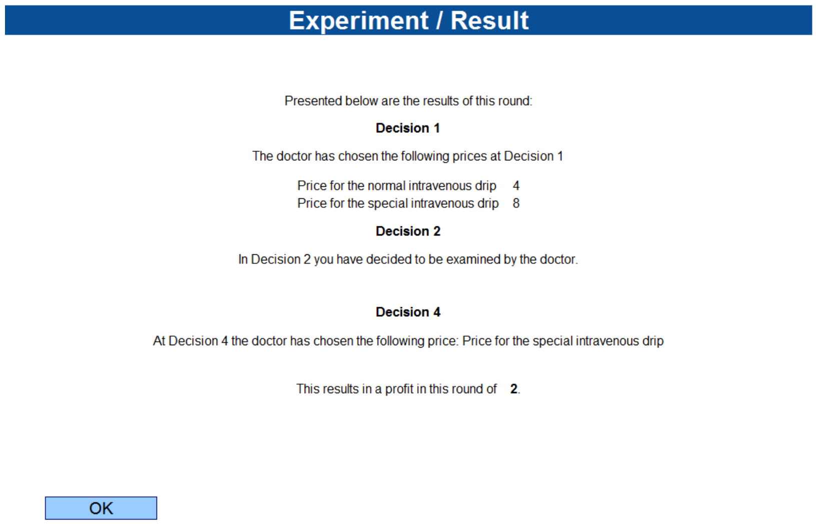


Figure D.2.30: Screenshot Patient / Result

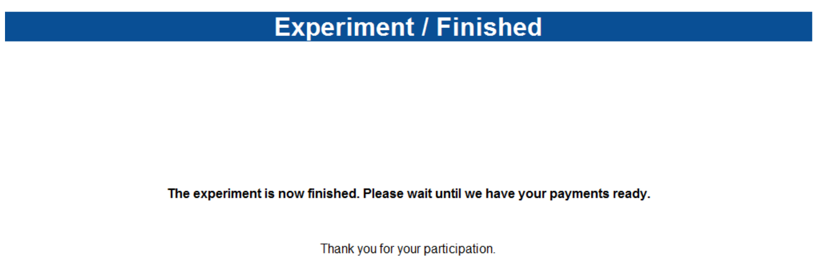


Figure D.2.31: Screenshot Doctor & Patient / Final Screen

D.2.4.2 Print Screen When Outside Option was Chosen

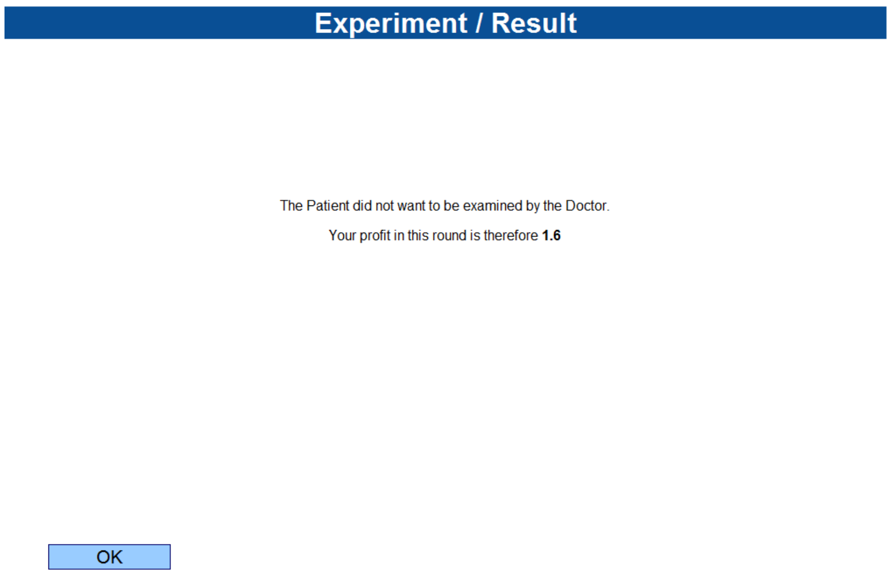


Figure D.2.32: Screenshot Doctor & Patient / Outside Option

D.2.5 Print Screens Verifiability Neutral

D.2.5.1 Player A & Player B

Experimental Instructions

Thank you very much for participating in this experiment. Please do not to talk to any other participant. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds** each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.
There are 2 roles in this experiment: **Player A** and **Player B. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.**

Player A always interacts with Player B. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 3 decisions, which are made sequentially. Decisions 1 and 3 are made by Player A, Decision 2 is made by Player B. At the start of the round Player B is randomly assigned a type (Type I or Type II).

Sequence of Decisions in a Round (in short)

1. Player A chooses a price for Action I and a price for Action II.
2. Player B is informed about the prices chosen by Player A. Player B then decides whether they want to interact with Player A.
3. Player A (but **not** Player B) is informed about the type of Player B. Then Player A is asked to choose either Action I or Action II. Player B has to pay the previously chosen price for the respective Action.
4. The results of this round, based on the decisions of Player A and Player B, will be announced.

Figure D.2.33: Screenshot Instruction Verifiability Neutral

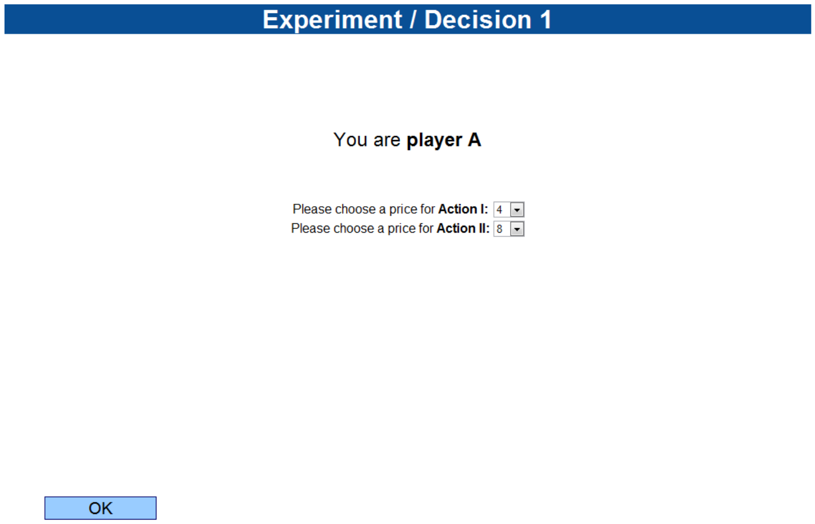


Figure D.2.34: Screenshot Player A / Decision 1

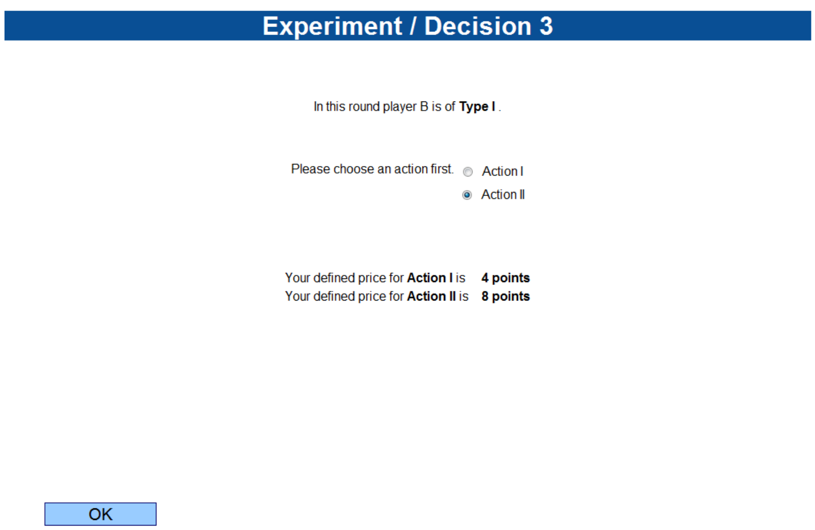


Figure D.2.35: Screenshot Player A / Decision 3 + 4

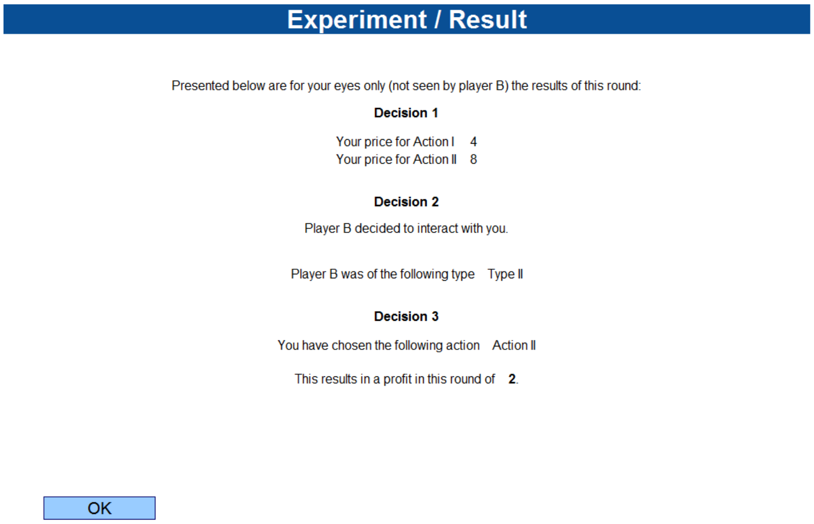


Figure D.2.36: Screenshot Player A / Result

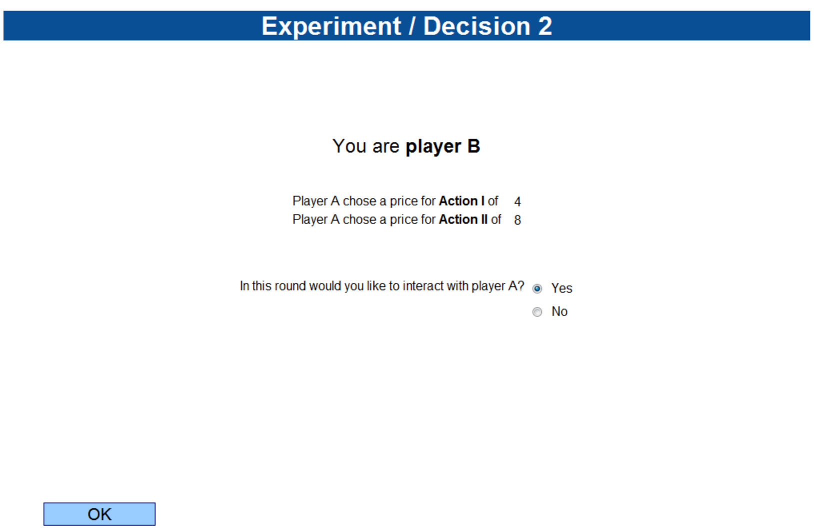


Figure D.2.37: Screenshot Player B / Decision 2

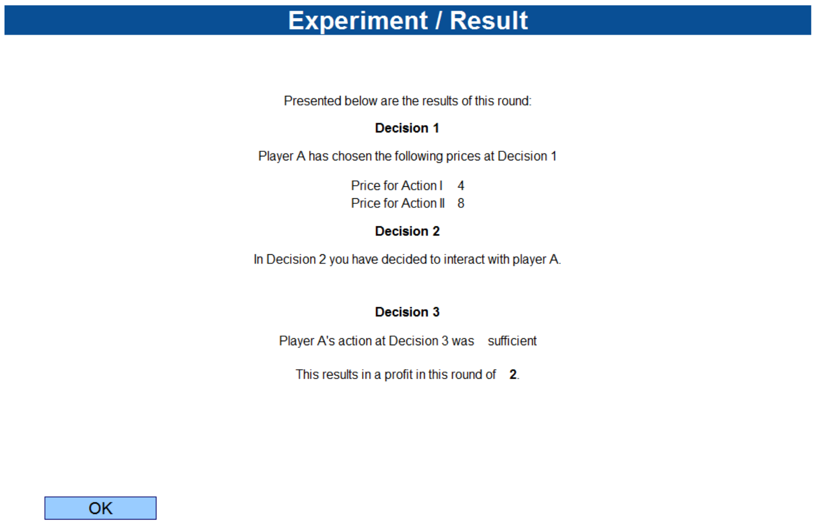


Figure D.2.38: Screenshot Player B / Result

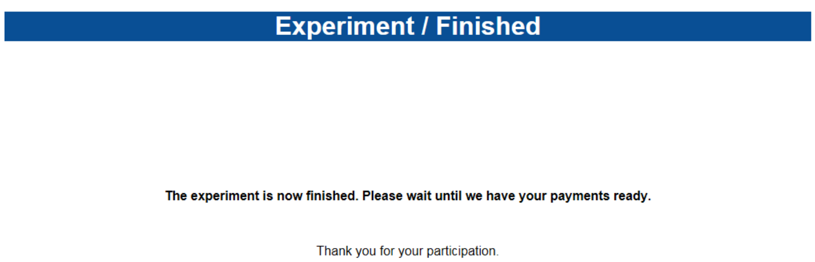


Figure D.2.39: Screenshot Player A & Player B / Final Screen

D.2.5.2 Print Screen When Outside Option was Chosen

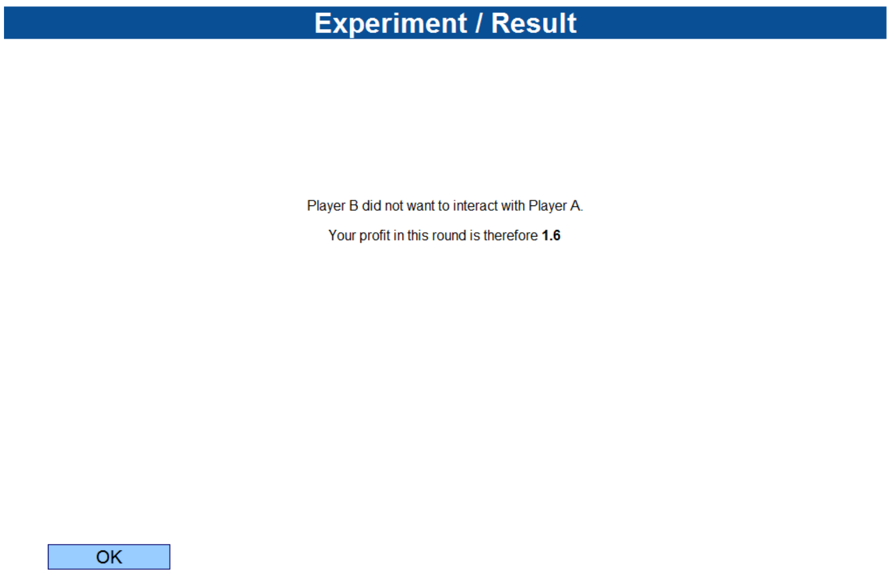


Figure D.2.40: Screenshot Player A & Player B / Outside Option

D.2.6 Print Screens Verifiability Medical

D.2.6.1 Doctor & Patient

Experimental Instructions

Thank you very much for participating in this experiment. Please do not to talk to any other participant until the end of the experiment. Please turn off your mobile phone or put it into flight mode.

2 Roles and 16 Rounds

This experiment consists of **16 Rounds** each of which has the same sequence of decisions. The sequence of decisions is explained in detail below.

There are 2 roles in this experiment: **Player A** and **Player B. At the beginning of the experiment you will be randomly assigned to one role and remain in that role throughout the experiment. On the first screen of the experiment you will see which role you have been assigned to.**

Player A always interacts with Player B. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

All participants receive exactly the same information regarding the rules of the game, including information about costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by Player A, Decision 2 is made by Player B. At the start of the round Player B is randomly assigned a type (Type I or Type II).

Sequence of Decisions in a Round (in short)

1. Player A chooses a price for Action I and a price for Action II.
2. Player B is informed about the prices chosen by Player A. Player B then decides whether they want to interact with Player A.
3. Player A (but **not** Player B) is informed about the type of Player B. Then Player A is asked to choose either Action I or Action II.
4. Player A charges one of the prices specified in Decision 1. The price charged **does not** necessarily have to be the price for the action actually chosen by Player A in Decision 3, it can also be the price of the other action.

Figure D.2.41: Screenshot Instruction Verifiability Medical

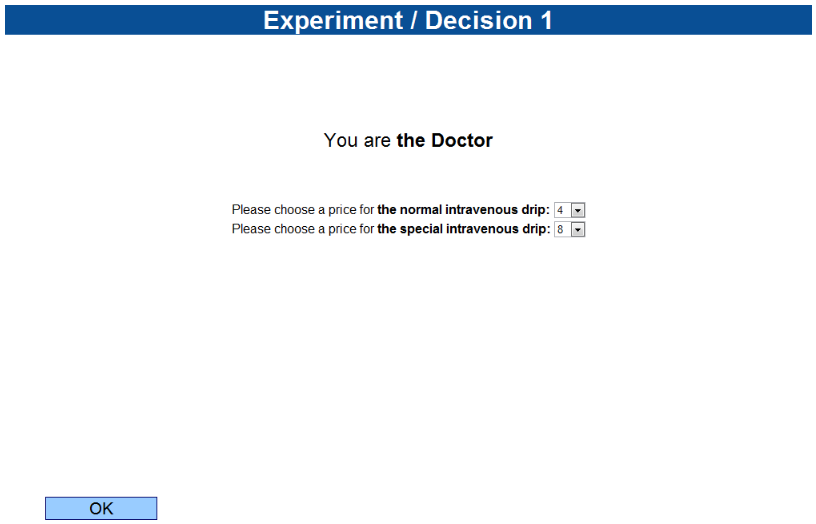


Figure D.2.42: Screenshot Doctor / Decision 1

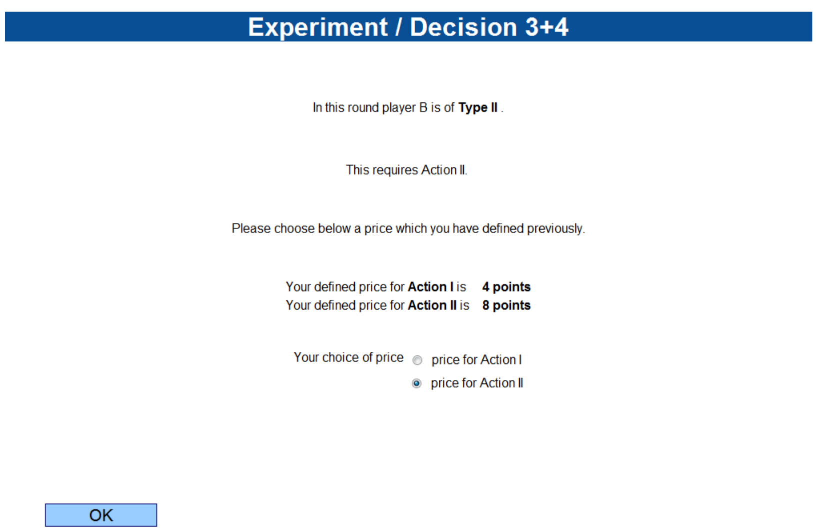


Figure D.2.43: Screenshot Doctor / Decision 3 + 4

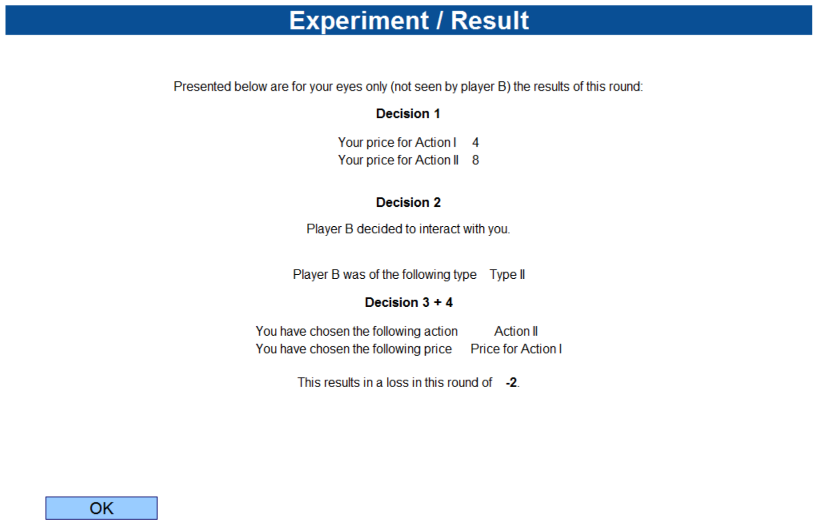


Figure D.2.44: Screenshot Doctor / Result

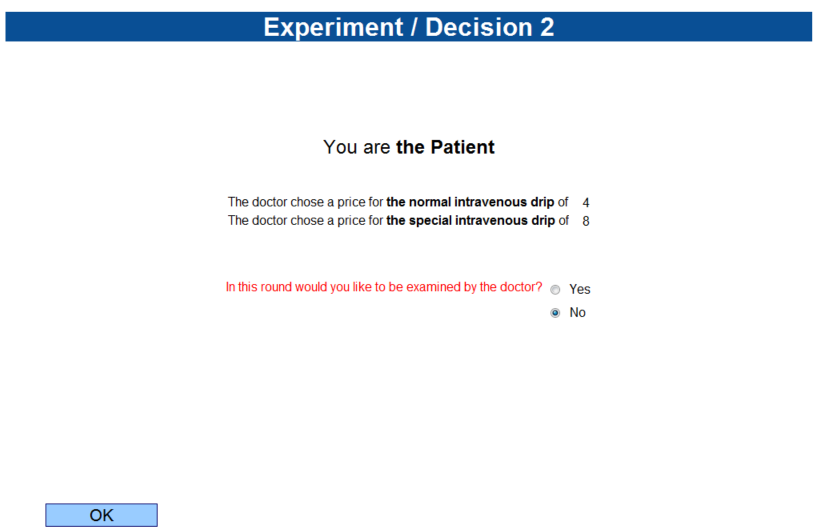


Figure D.2.45: Screenshot Patient / Decision 2

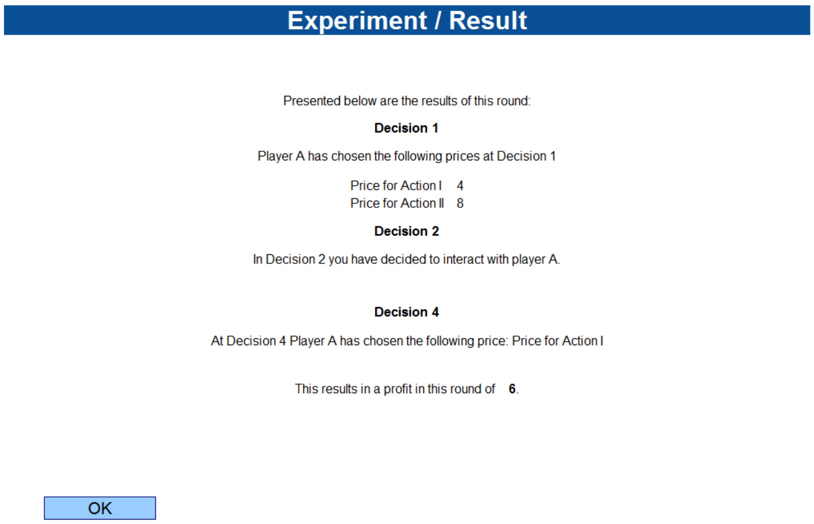


Figure D.2.46: Screenshot Patient / Result

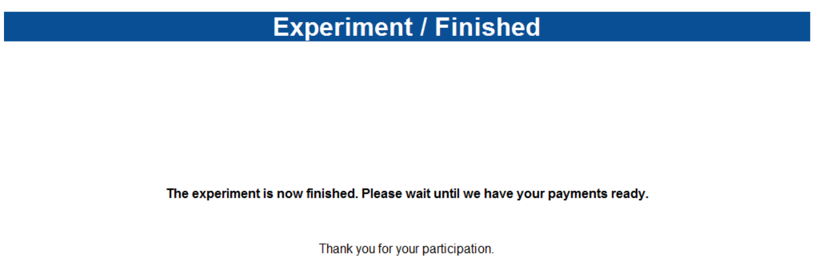


Figure D.2.47: Screenshot Doctor & Patient / Final Screen

D.2.6.2 Print Screen When Outside Option was Chosen

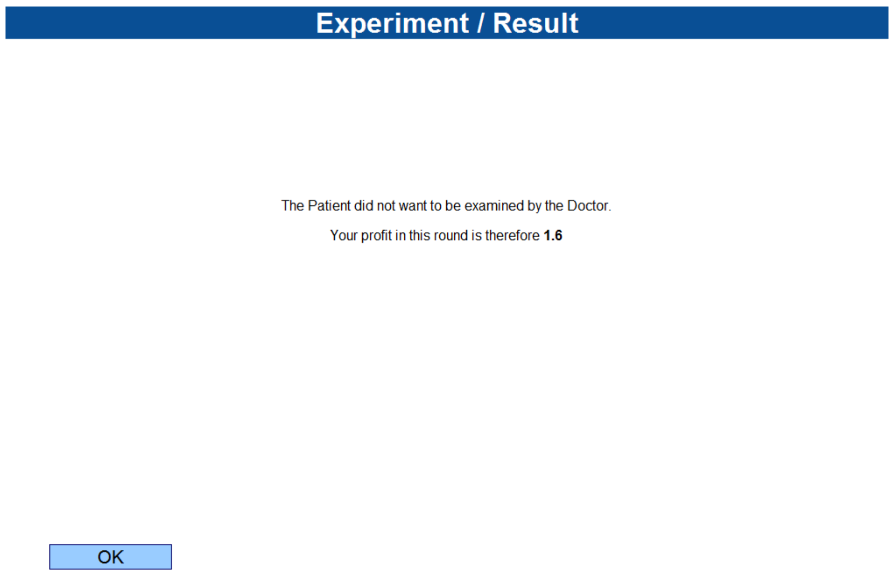


Figure D.2.48: Screenshot Doctor & Patient / Outside Option

D.2.7 Print Screens Triple Experiment

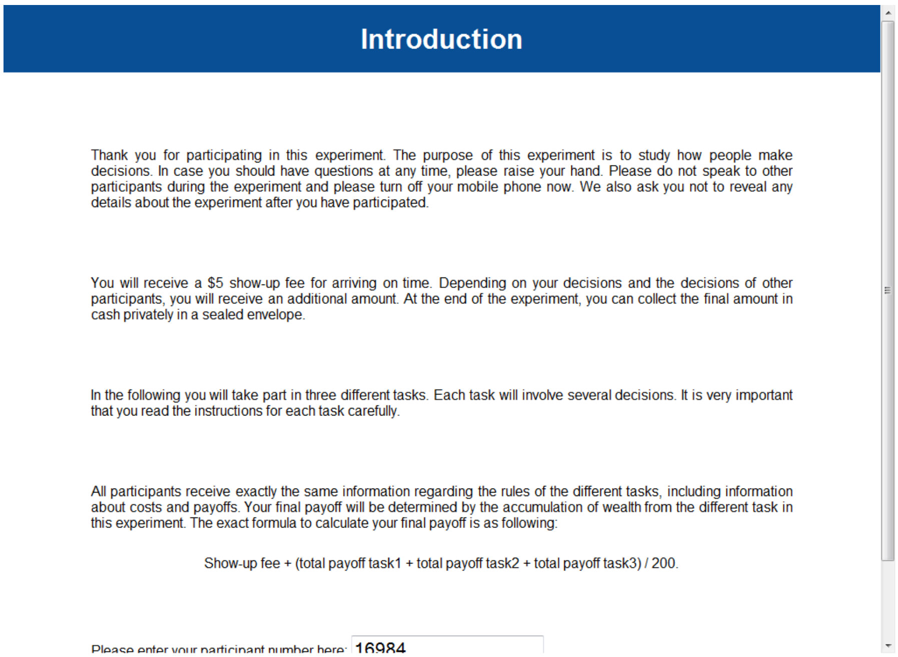
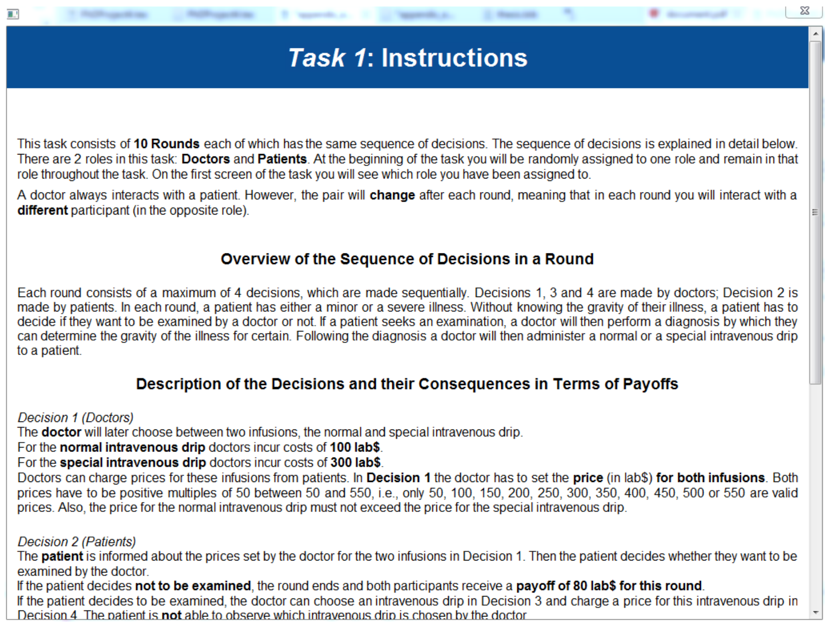


Figure D.2.49: Screenshot General Instruction



Task 1: Instructions

This task consists of **10 Rounds** each of which has the same sequence of decisions. The sequence of decisions is explained in detail below. There are 2 roles in this task: **Doctors** and **Patients**. At the beginning of the task you will be randomly assigned to one role and remain in that role throughout the task. On the first screen of the task you will see which role you have been assigned to.

A doctor always interacts with a patient. However, the pair will **change** after each round, meaning that in each round you will interact with a **different** participant (in the opposite role).

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 4 decisions, which are made sequentially. Decisions 1, 3 and 4 are made by doctors; Decision 2 is made by patients. In each round, a patient has either a minor or a severe illness. Without knowing the gravity of their illness, a patient has to decide if they want to be examined by a doctor or not. If a patient seeks an examination, a doctor will then perform a diagnosis by which they can determine the gravity of the illness for certain. Following the diagnosis a doctor will then administer a normal or a special intravenous drip to a patient.

Description of the Decisions and their Consequences in Terms of Payoffs

Decision 1 (Doctors)
The **doctor** will later choose between two infusions, the normal and special intravenous drip.
For the **normal intravenous drip** doctors incur costs of **100 lab$**.
For the **special intravenous drip** doctors incur costs of **300 lab$**.
Doctors can charge prices for these infusions from patients. In **Decision 1** the doctor has to set the **price** (in lab$) **for both infusions**. Both prices have to be positive multiples of 50 between 50 and 550, i.e., only 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 or 550 are valid prices. Also, the price for the normal intravenous drip must not exceed the price for the special intravenous drip.

Decision 2 (Patients)
The **patient** is informed about the prices set by the doctor for the two infusions in Decision 1. Then the patient decides whether they want to be examined by the doctor.
If the patient decides **not to be examined**, the round ends and both participants receive a **payoff of 80 lab$ for this round**.
If the patient decides to be examined, the doctor can choose an intravenous drip in Decision 3 and charge a price for this intravenous drip in Decision 4. The patient is **not** able to observe which intravenous drip is chosen by the doctor.

Figure D.2.50: Screenshot Task 1 - Instruction

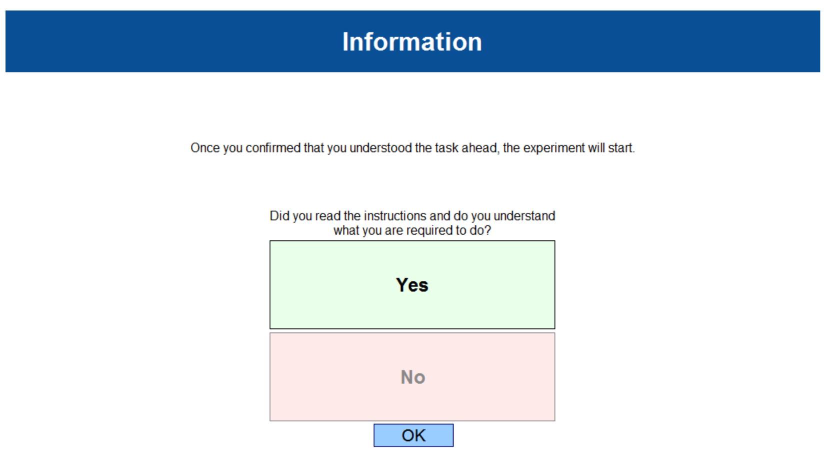


Figure D.2.51: Screenshot Understanding Instruction

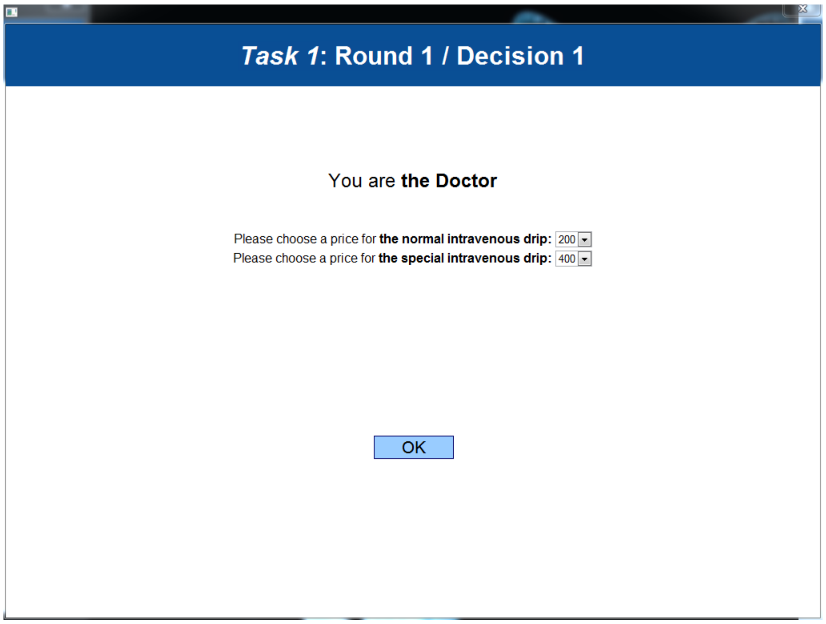


Figure D.2.52: Screenshot Task 1 - Credence Decision 1

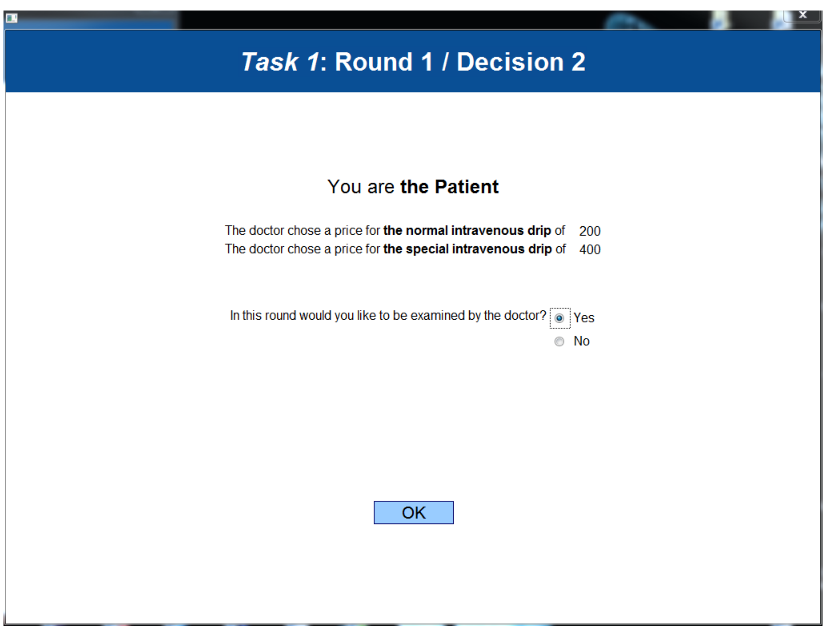


Figure D.2.53: Screenshot Task 1 - Credence Decision 2

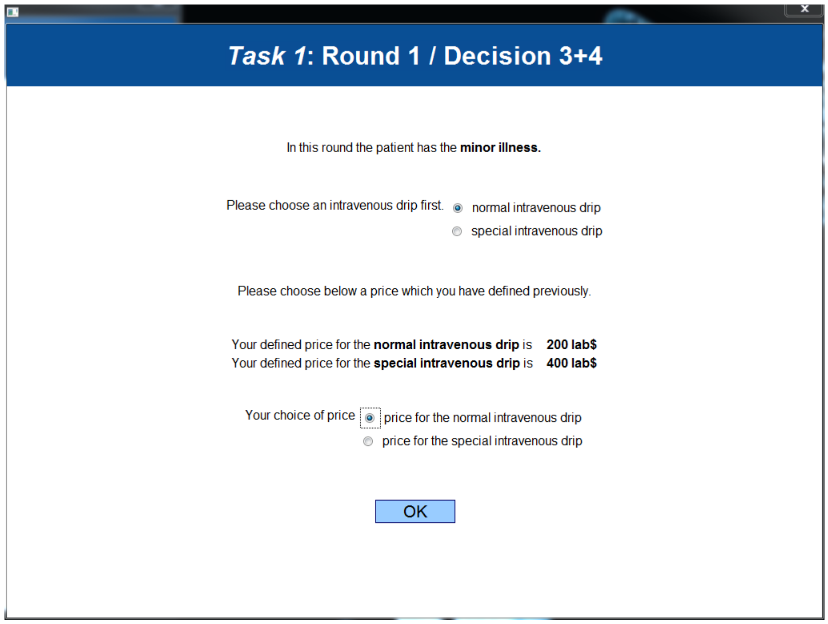


Figure D.2.54: Screenshot Task 1 - Credence Decision 3 + 4

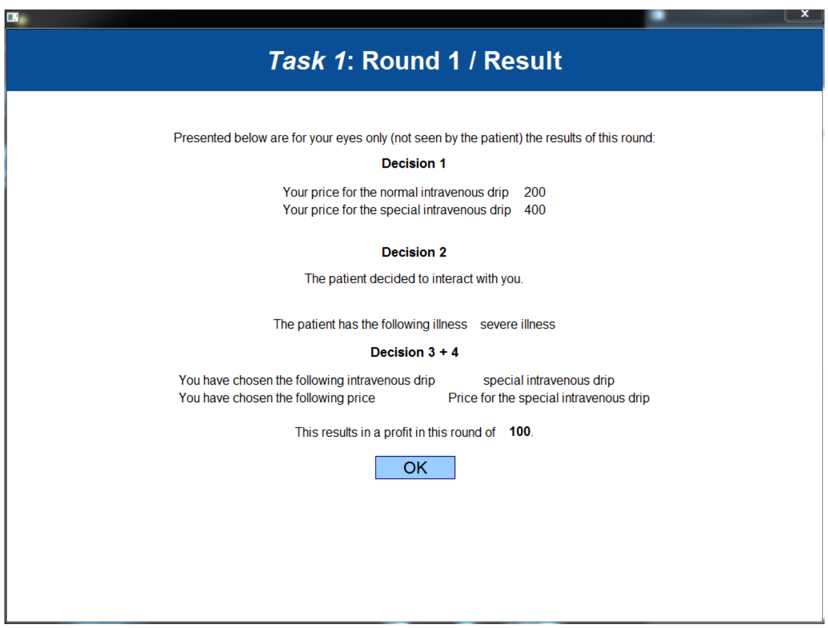


Figure D.2.55: Screenshot Task 1 - Credence Result Doctor

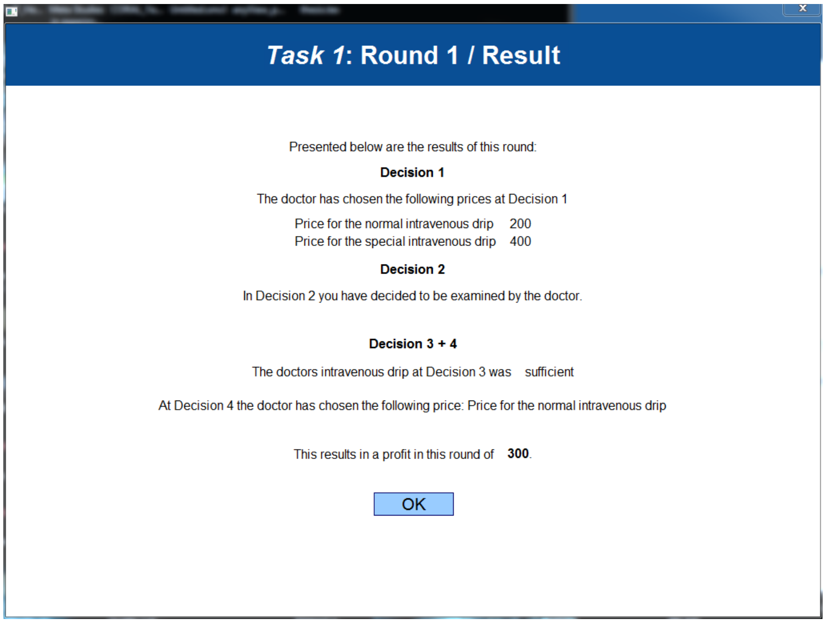


Figure D.2.56: Screenshot Task 1 - Credence Result Patient

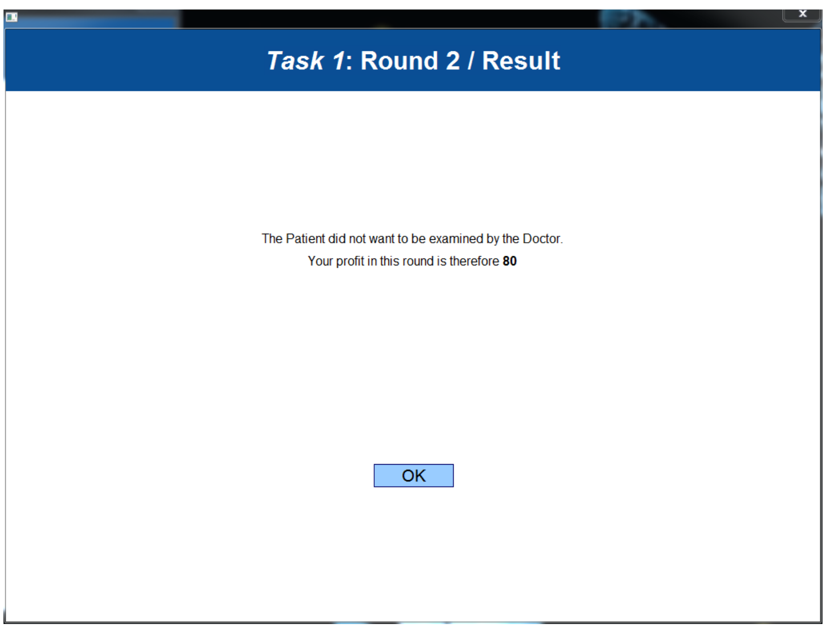


Figure D.2.57: Screenshot Task 1 - Credence Outside Option

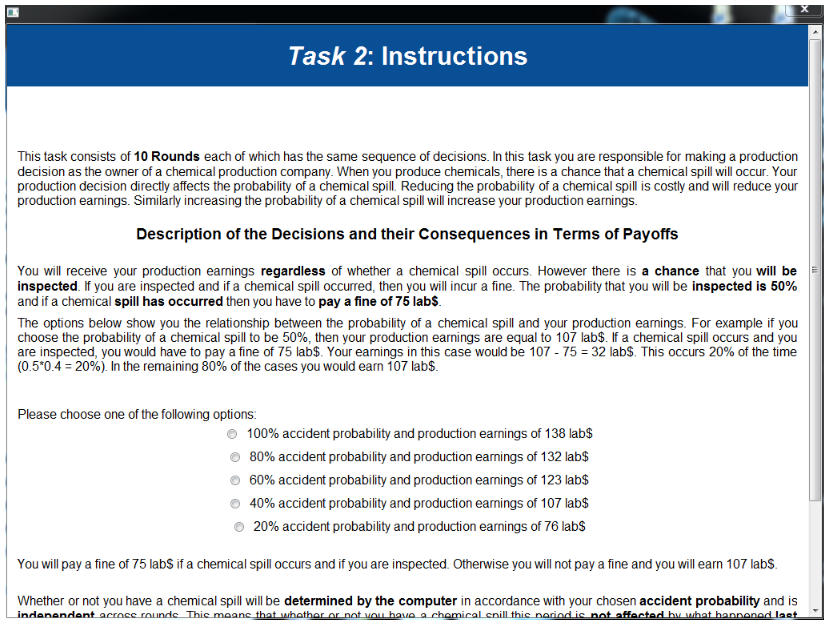


Figure D.2.58: Screenshot Task 2 - Self-Reporting Instruction

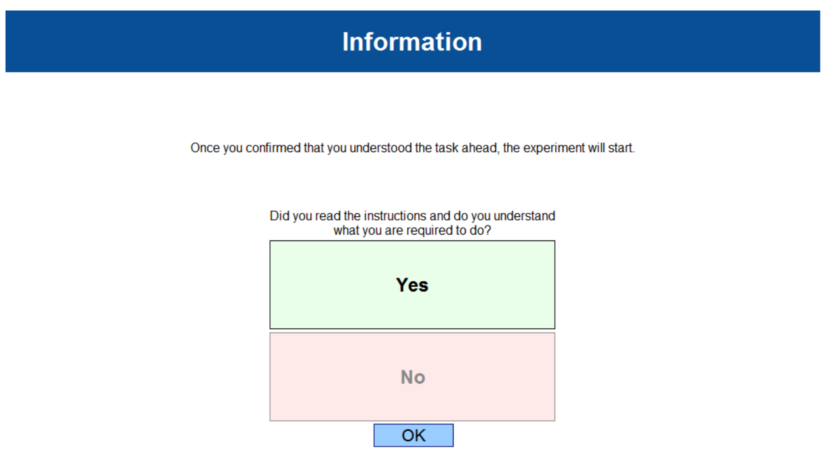


Figure D.2.59: Screenshot Understanding Instruction

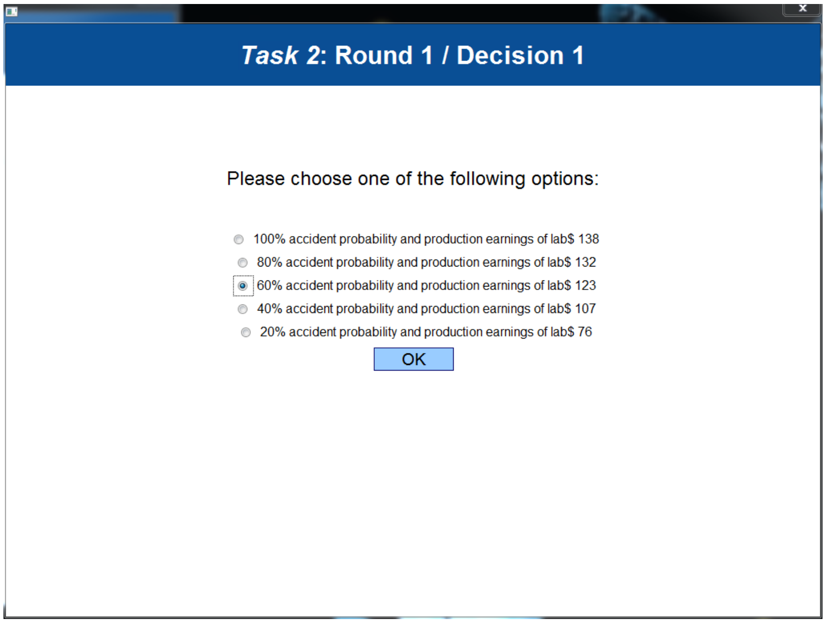


Figure D.2.60: Screenshot Self-Reporting Task - Production Decision

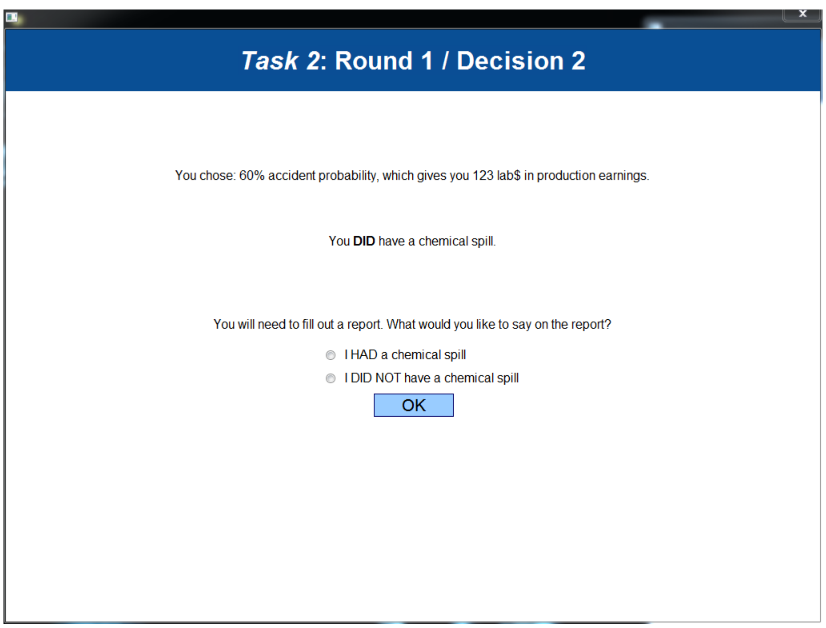


Figure D.2.61: Screenshot Self-Reporting Task - Did have Chemical Spill

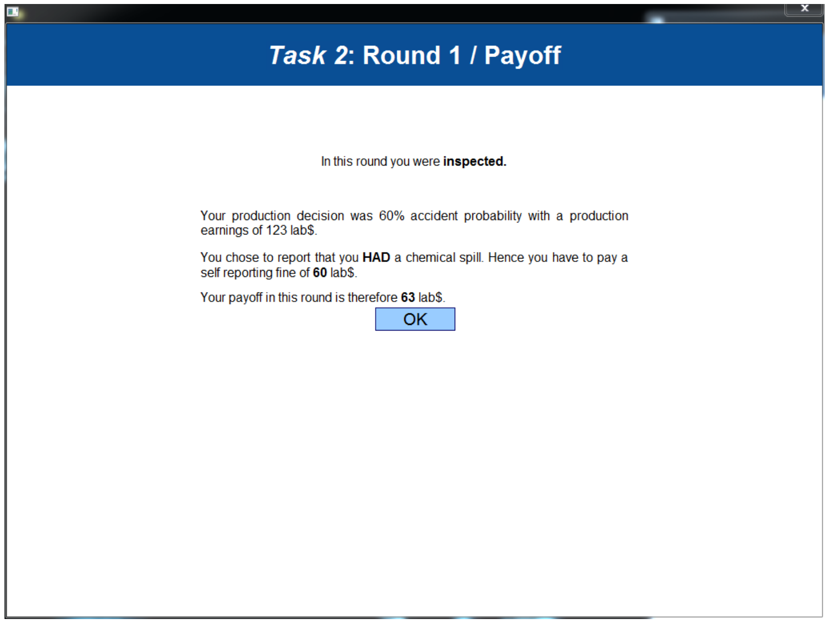


Figure D.2.62: Screenshot Self-Reporting Task - Payoff after Chemical Spill

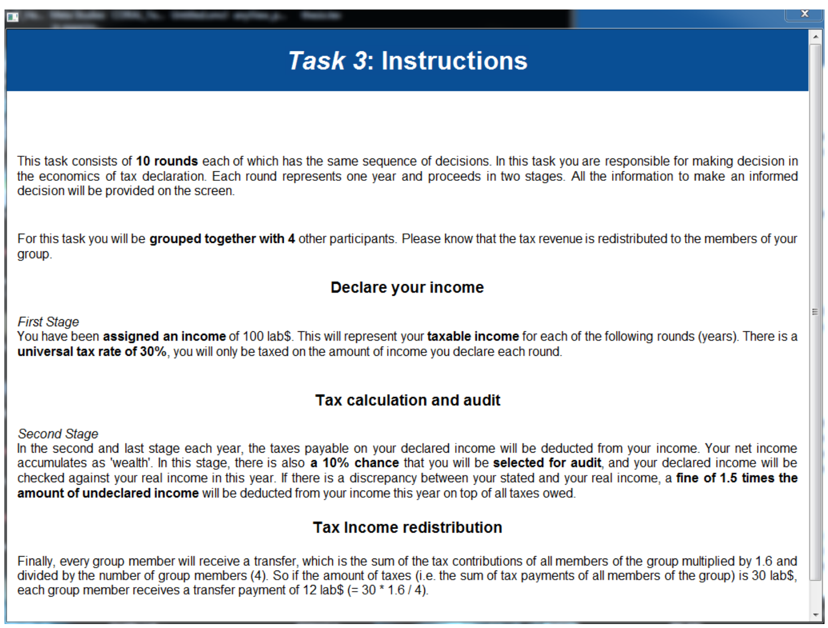


Figure D.2.63: Screenshot Task 3 - Tax Instruction

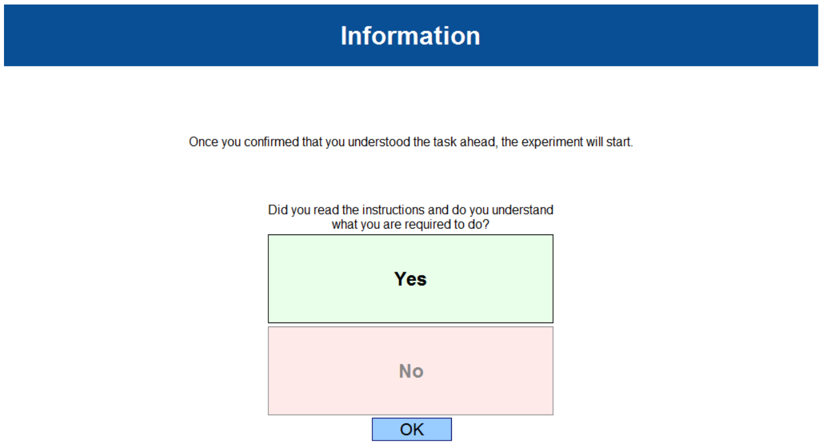


Figure D.2.64: Screenshot Understanding Instruction

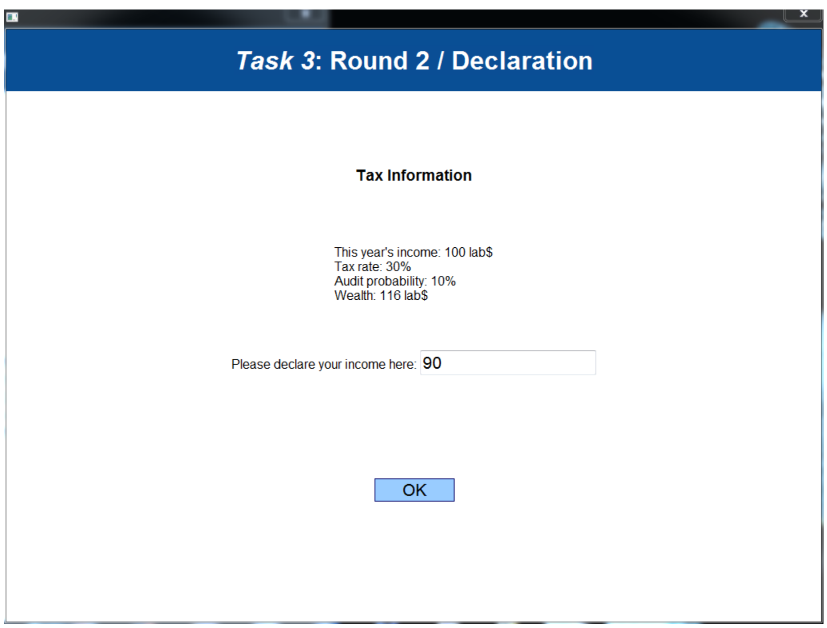


Figure D.2.65: Screenshot Task 3 - Tax Declaration

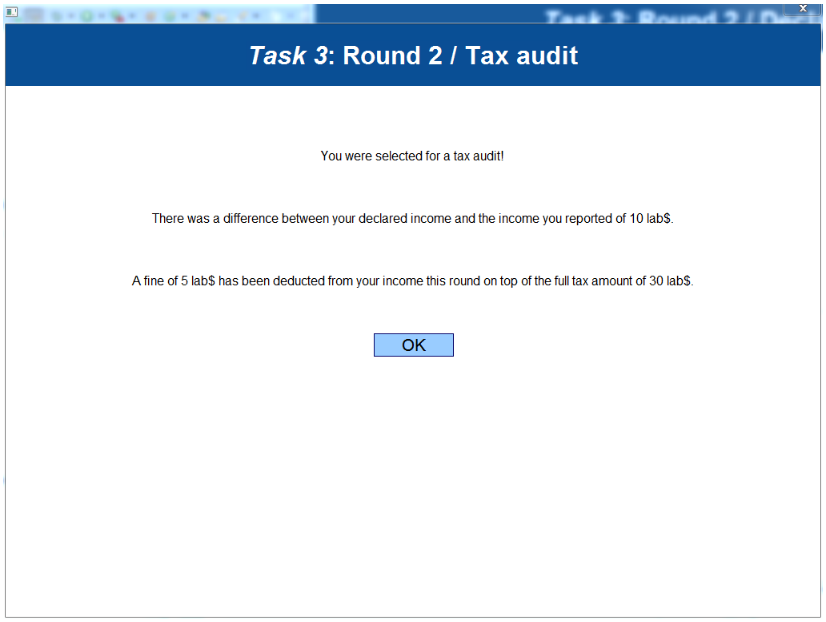


Figure D.2.66: Screenshot Task 3 - Tax Audit

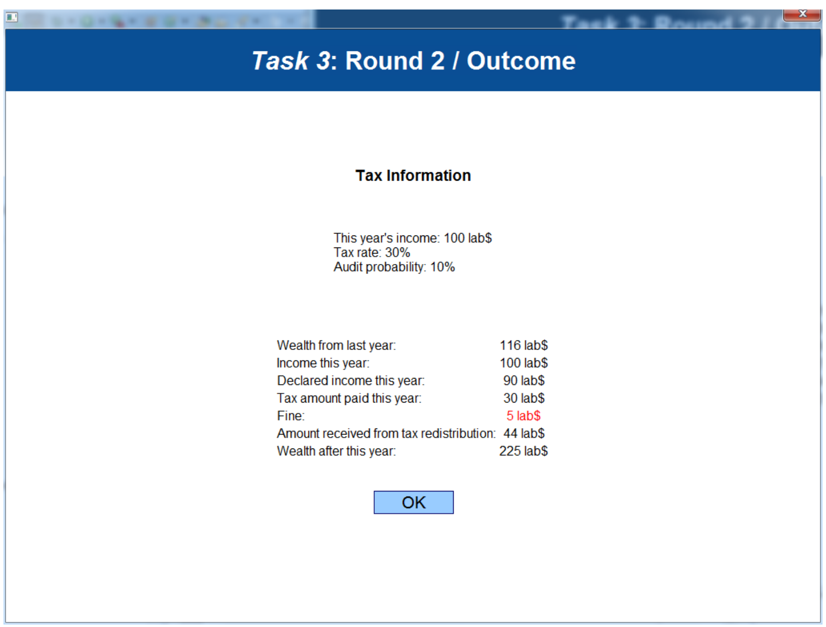


Figure D.2.67: Screenshot Task 3 - Tax Outcome